PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 16 January 15, 1930 Number 1

THE LIGHT-SENSITIVE SYSTEM AS THE BASIS OF THE PHOTIC RESPONSES OF PHYCOMYCES

By E. S. Castle

Laboratory of General Physiology, Harvard University

Communicated November 27, 1929

I.—Confusion regarding the nature of the steps intervening between the reception of light by phototropic plants and the resulting curvatures has led to controversy as to the validity of the simple theory proposed by Blaauw (1914; 1919), namely, that the phototropic bending of sessile plants or their parts is due to effective local inequalities in the rate of growth caused by corresponding differences in the amount or intensity of illumination. Blaauw specifically conceived the differences in rate of growth to be due to quantitatively different "light-growth" reactions, the latter being responses manifested as temporary accelerations or retardations of growth following sudden exposure to light of sufficiently high intensity. The indecisive character of recent work concerned with the cause of plant phototropism is made clear in the review by Brauner (1927).

The response of the light-sensitive sporangiophore of *Phycomyces* to sudden illumination of sufficient intensity is shown, after a distinct latent period during which the original growth rate is maintained, by a temporary acceleration in rate of growth. The reaction time is thus compound, consisting at least of an exposure period and a latent period. The duration of the latent period is constant for a particular intensity of light, unless the duration of the stimulating flash of light is reduced below a certain minimum. Below this minimum the reaction time lengthens progressively as the duration of the flash is shortened. It is possible to study the reaction time as a function of the duration of the exposure, using the direct acceleration of growth as the index of excitation.

The response of a sensitive sporangiophore to effective *unilateral* illumination consists, typically, in a so-called "positive" bending of the sporangiophore toward the source of light. This bending takes place close below the sporangium, precisely in the zone of normal, vertical growth. The phototropic response is, like the simple growth-response, compound, and

exhibits a comparable latent period. By using sufficiently brief times of exposure to unilateral light, the duration of this latent period also may be studied in the manner already indicated as a function of the duration of the exposure.

A new method, therefore, of examining the relation of the "light-growth" response and the phototropic response is to note the way a significant property of each, the latent period, changes as a function of a common variable, the duration of the exposure to light, under comparable conditions of experimentation. If each mode of response is similarly related to the common variable, it is reasonable to suppose that both have the same functional basis, namely, the light-sensitive system, the properties of which are being investigated.

II.—Pure cultures of *Phycomyces blakesleeanus* ("+" strain) were grown as previously described (Castle, 1927–28). For experimentation a culture was placed in a glass cell with plane walls and cover nearly submerged in a water thermostat enclosed by a dark room. Directly above the sporangiophore under consideration was a 100-watt incandescent bulb serving to orient the sporangiophore precisely and at the same time adapt it to a particular illumination (86 ft.-candles). Inside the glass cell, beside the culture, stood a 45° mirror which reflected laterally onto the sporangiophore the stimulating beam of light. This light descended vertically onto the mirror, through a shutter and a colloidal-gold heat-screen, from a 1000-watt lamp enclosed above in a special housing. At the sporangiophore the intensity of this light was 171 ft.-candles. Observation of the sporangiophore was made laterally against an intermittent red light by means of a horizontal microscope projecting into the dark room and containing an ocular micrometer scale. By turning this ocular scale into either a vertical or a horizontal position it could be used to observe, respectively, the vertical growth or the lateral bending of the sporangiophore.

In the actual procedure—(1) a sporangiophore was adapted to the orienting light of 86 ft.-candles for 30 minutes; (2) the orienting light was then put out, and dark adaptation of 30 minutes' duration allowed to take place; (3) stimulation with unilateral illumination of 171 ft.-candles was effected for varied brief times of exposure by means of the 1000-watt lamp and properly calibrated shutters; (4) the interval between the beginning of stimulation and the subsequent growth-reaction was determined from plottings of the position of the sporangium on the micrometer scale at 15 second intervals (cf. Castle, 1928–29).

The reaction time is taken as the interval between the beginning of stimulation and the first perceptible growth-reaction, either acceleration of growth or bending. It must be pointed out that a normal, vertical growth acceleration typically follows stimulation of the sporangiophore

with unilateral light, as a consequence of its great transparency, even though there is also a practically simultaneous bending. The experimental conditions, therefore, under which acceleration of growth and bending are here compared are identical. The *beginning* of either response

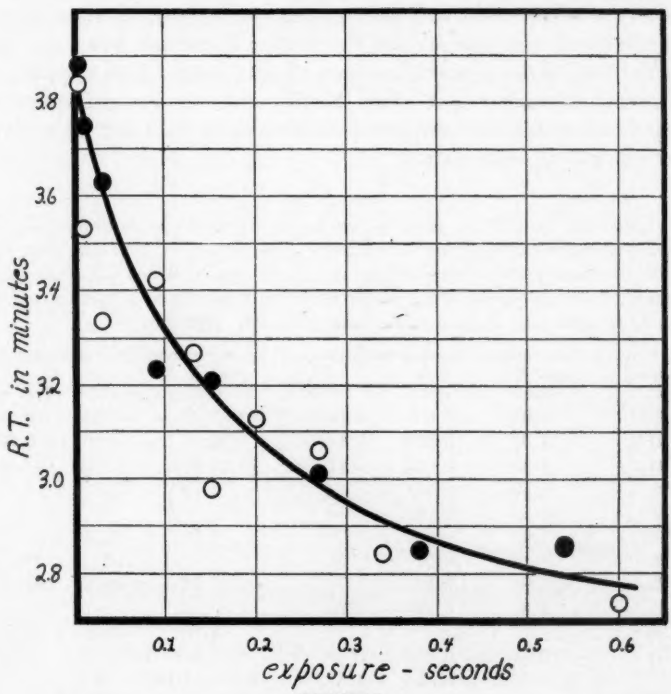


FIGURE 1

Reaction times of the sporangiophores of *Phycomyces* as a function of the duration of exposure to light. The open circles are the data for the direct growth-response, the solid circles those for phototrophic response. To the latter 0.25 min. has been added in this plot, for the reason explained in the text. Each point is the average of from 13 to 22 individual determinations. The curve is a hyperbola, one constant of which is taken from the slope of the line in figure 2.

is the measured end-point, and the position of the ocular scale, whether vertical or horizontal, determines which response shall be observed.

III.—When sporangiophores under comparable conditions of high sensitivity are stimulated at constant temperature by unilateral light of constant intensity but of varied duration, a regular relationship is found to hold between the reaction time and the duration of the exposure to

light, for both growth acceleration and phototropic bending. The two sets of data in table 1 are plotted together in figure 1. The same curve may evidently be made to fit both, provided a constant time (15 seconds) is added to the reaction times for the phototropic response. This constant is the interval between successive readings of the position of the sporangium on the ocular scale, and the fact that the reaction time figures from the phototropic response are, on the whole, 15 seconds less than those from the other series is merely evidence that as judged from the plots the phototropic response begins more sharply, and the first perceptible response is estimated one unit (15 seconds) earlier than the more slowly occurring growth acceleration.

TABLE 1

Mean reaction times of the direct growth-response and of the phototropic response to various durations of exposure to unilateral illumination of 171 ft.-candles. Each mean *R.T.* represents the average of from 13 to 22 determinations on individual sporangiophores. Such averageing of the photic responses is justifiable in spite of the variability in absolute rate of growth (cf. Castle, 1927–28; 1928–29).

	DIRECT GROWTH-RESPONSE			PHOTOTROPIC RESPONSE		
EXPOSURE (SEC.)	MEAN *R.T.* (MIN.)	*P.E.* OF MEAN *R.T.*	$\frac{1}{R.T. - 2.50}$	MEAN *R.T.* (MIN.)	*P.E.* OF MEAN *R.T.*	$\frac{1}{R.T. - 2.25}$
0.005	3.84	±0.035	0.746	3.63	±0.028	0.725
0.01	3.53	0.029	0.971	3.50	0.061	0.800
0.03	3.33	0.042	1.21	3.38	0.040	0.885
0.09	3.42	0.031	1.09	2.98	0.024	1.37
0.13	3.27	0.041	1.30	..	...	...
0.15	2.98	0.046	2.08	2.96	0.030	1.41
0.20	3.13	0.036	1.59	..	...	...
0.27	3.06	0.057	1.79	2.76	0.037	1.96
0.34	2.84	0.030	2.94	..	...	...
0.38	..	...	...	2.60	0.048	2.86
0.54	2.86	0.045	2.79	2.61	0.028	2.79
0.6	2.7	0.02	4.17	..	...	...

The comparability of the two series of data, one of them corrected as just explained, may be shown in another way, that is, by plotting the reciprocal of the reaction time, the latter minus the apparent "action time" of the response, against the duration of the exposure to light, or $1/(R.T. - M)$ against t (Fig. 2). M, the "action time," may be mechanical, but is of unknown nature. It is at any rate a time during which processes occur distinct from those of the exposure time and of the latent period proper, the latter being construed as resulting directly from photochemical action. By choosing a suitable value for M (2.50 min.), a linear relationship between $1/(R.T. - M)$ and t can be established, holding equally for both of the modes of response. It is interesting that this linear relationship is similar to that which has been found for the duration of the latent period in the response of the clams *Mya* and *Pholas* to light (Hecht,

1918–19; 1927–28) and for the latent period in the electrical response of the eel's retina to light (Adrain and Matthews, 1927). The relation may be interpreted in a number of ways, but certainly indicates that over the range of exposure times studied in *Phycomyces*, the velocity of the process occurring during the latent period is directly proportional to the amount of preceding photochemical action.

Summary.—(1) The reaction time of the direct growth-response of the sporangiophore of *Phycomyces* to light consists of a series of at least

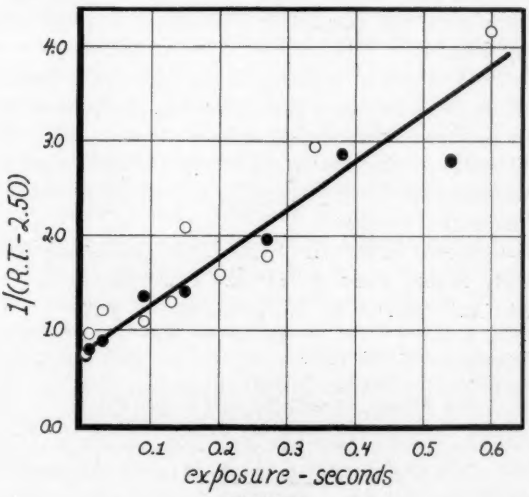


FIGURE 2

Reciprocal of reaction time (the latter minus 2.50 min., the apparent "action-time" of the response) plotted against duration of exposure. Symbols as in figure 1. The phototropic response is corrected for its consistently earlier appearance by plotting $1/(R.T.-2.25)$ instead of $1/(R.T.-2.50)$. Comparison of figure 1 and of the probable errors in table 1 will show that the deviations of the points in this plot from the straight line are not unreasonably great.

three major components: (*a*) an exposure period during which photochemical change occurs; (*b*) a "latent period" involving products directly consequent upon the photochemical action; and (*c*) an action-time occupying a further interval before the growth acceleration appears.

(2) The reaction time of the phototropic response of the sporangiophore following stimulation by unilateral illumination is also compound, and is made up of at least three components similar to those of the direct growth-response.

(3) The reaction time of the direct growth-response is constant for a particular intensity of illumination, provided that the duration of the exposure period exceeds a certain value. Below that value the reaction time increases as the exposure time decreases. This is also true for the phototropic response.

(4) The manner in which the reaction time of each mode of response varies as a function of the time of exposure to light is shown to be similar. It is therefore concluded that the two responses have the same functional basis, namely, the light-sensitive system under investigation. This conclusion accords with the reasonable conception of plant phototropism developed by Blaauw.

(5) With both modes of response, if an appropriate constant is subtracted from the reaction times the reciprocals of the resulting numbers follow a linear sequence when plotted against the durations of the exposure to light. The rate of the process occurring during the latent period is therefore considered to be directly proportional to the amount of preceding photochemical change.

Part of the expense incidental to this investigation was defrayed by a grant from the Milton Fund of Harvard University.

Adrian, E. D., and Matthews, R., *J. Physiol.*, **64**, 279, 1927.

Blaauw, A. H., *Zeitschr. f. Bot.*, **6**, 641, 1914; *Mededeelingen van de Lahdbouw hogeschool, Wageningen*, **15**, 89, 1919.

Brauner, L., *Ergebnisse Biol.*, **2**, 95, 1927.

Castle, E. S., *J. Gen. Physiol.*, **11**, 407, 1927–28; *J. Gen. Physiol.*, **12**, 391, 1928–29.

Hecht, S., *J. Gen. Physiol.*, **1**, 657, 1918–19; *J. Gen. Physiol.*, **11**, 657, 1927–28.

THE RATE OF DECOMPOSITION OF NITROGEN PENTOXIDE AT VERY LOW PRESSURES

BY H. C. RAMSPERGER AND R. C. TOLMAN

GATES CHEMICAL LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated December 16, 1929

§1. *Introduction.*—In a previous article[1] the results have been reported of some experiments on the rate of decomposition of gaseous nitrogen pentoxide at moderately low pressures in the range 0.2 to 2 mm. of mercury. The special features of the experimental method employed, included the use of a very large reaction vessel constructed from a 45-liter flask, thus minimizing the effect of the walls, and the use of a specially sensitive "click" gauge to follow the reaction by direct pressure measurements, thus avoiding the dangers involved in following the reaction by periodically freezing out the oxides of nitrogen and measuring the evolved

oxygen, as has sometimes been done. The results obtained at these moderately low pressures showed no falling-off in the specific unimolecular rate of decomposition below that obtained at high pressures.

In the present article we desire to report the results of some further experiments on the rate of decomposition of nitrogen pentoxide at total initial pressures down to 0.0055 mm.

§2. *The Apparatus.*—Except for the pressure-measuring device, the apparatus was essentially the same as that previously described. To follow the change in pressure we adopted the general form of the capacitance gauge devised by Olson and Hirst.[2] In this gauge, the gas whose pressure is to be measured presses against a thin glass diaphragm to which is attached one of the plates of an electrical condenser, the other plate having a fixed position. This condenser is connected in series with a variable condenser, the two condensers together forming the capacity of an oscillating circuit which is loosely coupled with a neighboring circuit, having a constant frequency of oscillation maintained with the help of a piezo-electric quartz crystal. Changes in pressure are then measured by the reading of the variable condenser necessary to secure resonance.

In the final form in which we used this gauge, the movable diaphragm was made from a thin sheet of mica lesst han 0.002 in. thick, in order to increase sensitivity. This diaphragm was 2 in. in diameter and was sealed at its circumference with "Duco" lacquer so as to close the top of a Pyrex glass chamber which was shaped like an inverted bell and connected at its lower end to a tube leading to the reaction flask.

The movable condenser plate was also made from a thin sheet of mica of approximately the same diameter as the diaphragm and made electrically conducting by gold plating with a commercial preparation used in china painting. It was placed parallel to the diaphragm and cemented at its center to a small glass spacer, cemented in turn to the center of the diaphragm. The electrical connections with this movable condenser plate were made through a narrow strip of thin tin foil so arranged as not to put any strain on the motion of the diaphragm.

The fixed plate of the condenser was made of brass. It was supported in position parallel to the movable plate by a second bell-shaped Pyrex glass chamber ring-sealed to that supporting the diaphragm, in order that temperature changes should not affect the relative position of the two condenser plates, since they were then both supported by practically the same length of Pyrex glass. The distance between the two condenser plates was about 0.3 to 0.4 mm.

Stopcocks were introduced between the gauge and the reaction vessel so that the space below the diaphragm could be opened to the reaction vessel, filled with air at any desired pressure, or completely pumped out. It was also arranged so that we could have vacuum or any desired air

pressure over the diaphragm. The air pressure on either side of the diaphragm could be measured with the help of a McLeod gauge, liquid air traps being introduced between the McLeod gauge and the capacitance gauge so that the vapor pressure of mercury was not registered on the capacitance gauge. These arrangements were made so that the gauge could be continuously calibrated throughout the individual runs.

The electrical circuits for the measurement of capacity had the same layout as given by Olson and Hirst. Except for the A and B batteries, all parts of the circuit, together with the gauge itself, were enclosed in an air thermostat provided with a regulator. During the course of a run, slight adjustments in temperature were purposely made from time to time in order to compensate for a drift in the zero that otherwise occurred. The different portions of the electrical circuits were carefully fastened in position to prevent changes in their relative positions, and electrical shielding was introduced between the circuits and the observer.

We desire to express our thanks to Dr. Arnold Beckman and Dr. Martin Nordberg for assistance in the construction of this gauge.

§3. *The Experimental Procedure.*—The runs were made in three sets. The first set consisting of runs Nos. 1 and 2 was terminated by an accident to the gauge which necessitated a new diaphragm. The second set consisted of runs Nos. 2 to 7.

Following these, a third set consisting of runs Nos. 8 to 11 was made to determine the effect of wall surface. This was done by filling the 45-liter flask with 56 pieces of previously unused soft glass tubing, having an outside diameter of approximately 2.4 cm., and ranging in length from 20 to 45 cm. The tubes were cleaned by wiping on the outside with a dry towel, and on the inside with dry waste followed by very slightly dampened waste, and this in turn again by dry waste. These tubes filled the flask practically full, increasing the ratio of surface to volume from about 0.14 $cm.^{-1}$ to about 0.73 $cm.^{-1}$, and very greatly decreasing the distance that an average molecule has to travel to come in contact with the surface.

Before each set of runs, the flask was pumped down to a pressure of about 2×10^{-5} mm. and heated to about 70° to 80°C. for a period of 15 to 20 hours, nitrogen pentoxide being admitted, allowed to decompose, and then repumped out, once or twice during this period to assist in conditioning the surface. The inside of the reaction vessel was at all times protected from mercury vapor by liquid air traps.

The runs were made at approximately 40°C. The temperature was held constant throughout each run to within 0.2°C. and usually much closer, but showed by soundings at 8 different places in the air thermostat a maximum variation of 0.8°C. between different parts of the thermostat. This variation was allowed for by averaging and the final figures given are believed to be reliable to about 0.2°C.

The pentoxide was prepared as in the previous work, and stored in crystalline form in a cooled fore-vessel from which it could be admitted to the reaction flask by evaporation. Working at the very low pressures used in this investigation, the final pressure obtained at the end of a run was never as large as would be calculated from the initial pressure on the assumption that we had pure pentoxide at the start. Such a lack of purity may have resulted from an accumulation of decomposition product in the surface layers of the pentoxide.

In making the individual runs, the rate of decomposition of the pentoxide was followed by pressure measurements made with the capacitance gauge described above. To do this, the gauge was first pumped out and a zero reading for the capacity obtained; connection was then immediately made with the reaction vessel and a new reading for the capacity made. In addition spaced throughout the run an approximately equal number of similar pairs of readings were taken with air pressure on the gauge, these pressures being then measured with a McLeod gauge. In this way the gauge was continuously calibrated against known pressures. The amount of reaction mixture removed by making the pressure measurements in this way was not serious since the reaction flask had a volume of 45,000 cc. and the gauge chamber which had to be pumped out had a volume of only 30 cc.

In obtaining the capacity measurements the procedure which it was found best to adopt was to start with such a capacity in the gauge circuit that the crystal was not oscillating, and then increase the capacity to a point where the crystal started oscillating, which was marked by a sharp decrease in plate current. When readings were not being taken the crystal was always kept not oscillating to prevent heating. In taking the readings several settings were always made in succession, since there was a tendency for a slight drift (several scale divisions) in the readings until a constant value was obtained.

The sensitivity of the pressure-measuring device depended on a number of factors, including the diaphragm used, the total pressure being measured, and the ratio between the capacity of the gauge and the variable capacity in the secondary circuit. For these reasons it varied from about 3.3×10^{-5} mm. to 13.7×10^{-5} mm. for one scale division on the variable condenser (General Radio Co., Type 222). The computed pressures appeared to be reliable to about 1 or 2%.

§4. *The Experimental Results.*—In order to give an idea of the reliability of the results we give, as examples, in tables 1 and 2 the data obtained in run No. 1 where the pressure was moderately high and the specific reaction rates checked well, and in run No. 7, which was our lowest pressure run and gave a relatively large deviation in the individual values for specific reaction rate.

The specific rates were calculated on the basis of a first order decomposition by the interval method, since it gives a fairer idea of any trend in the reaction, even though it gives a poorer check among constants than the overall method. No evidence of trend in rate was found in the case of any of the runs made in the flask without the tubes, but there was always a dropping-off in the calculated specific rates with the progress of the run after the surface had been increased by the introduction of the tubing. This dropping-off in the case of the additional surface was least at the highest pressure.

TABLE 1

Run No. 1 Temperature = 40.04°C.

TIME SECONDS	TOTAL PRESSURE MM.	PRESSURE OF N_2O_5 MM.	SPECIFIC REACTION RATE k_1 (SEC.$^{-1}$) $\times 10^4$
0			
30	0.0431	0.0370	
			2.53
330	0.0472	0.0343	
			2.32
880	0.0534	0.0301	
			2.48
1650	0.0612	0.0249	
			2.52
2520	0.0686	0.0200	
			2.44
3435	0.0746	0.0160	
			2.21
4410	0.0796	0.0129	
			2.84
5450	0.0842	0.0096	
16940	0.0972	0.0010	
∞	0.0986	0.0000	
		Average	2.48
		Mean deviation	0.13

TABLE 2

Run No. 7 Temperature = 40.32°C.

TIME SECONDS	TOTAL PRESSURE MM.	PRESSURE OF N_2O_5 MM.	SPECIFIC REACTION RATE k_1 (SEC.$^{-1}$) $\times 10^4$
0			
100	0.00565	0.00443	
			2.36
655	0.00645		
1310	0.00730	0.00333	
			1.50
1950	0.00770		
2735	0.00827	0.00269	
			1.86
3610	0.00874		
4895	0.00960	0.00180	
			1.77
6190	0.00995		
7680	0.01065	0.00110	
12260	0.01168		
∞	0.01230		
		Average	1.87
		Mean deviation	0.24

The specific rates were calculated for intervals so chosen as to give a fair picture of the course of the reaction. This was done by dividing the run into a reasonable number of intervals (4–7) in which approximately equal increases in pressure occurred. The first pressure reading, taken as

soon as possible ($^1/_2$ to 2 min.) after admission of the gas to the flask, was always used as the starting point for the calculations. The initial pressure of the nitrogen pentoxide was calculated by assuming complete decomposition at the end of the run.

A summary of the results obtained in all 11 runs is given in table 3 which is self-explanatory. The runs are arranged in the order of decreasing total initial pressures. The last column gives the ratio of the average specific rate of the run to that calculated for high pressures for the same temperature from the results of Daniels and Johnston.

§5. *Discussion of the Results.*—It is our belief that the results of runs Nos. 1 to 7, as summarized in table 3, give a reasonably correct idea of the homogeneous rate of decomposition of nitrogen pentoxide at very low pressures. In accordance with our previous work, mentioned above, no falling-off in the specific rate was found down to pressures of about 0.2 mm., but in accordance with the present results a falling-off becomes evident at a total initial pressure around 0.04 mm. and continues to become more pronounced down to our lowest pressures around 0.006 mm., where the rate has dropped to about $^2/_3$ of its high pressure value. Throughout the whole pressure range, however, the rate of homogeneous decomposition is essentially first order, with no trend in the calculated constants with the progress of the run.

TABLE 3

RUN. NO.	TEMP. °C.	TOTAL INITIAL PRESSURE MM.	INITIAL PRESSURE N_2O_5 MM.	NO. CONSTANTS CALCULATED	SPECIFIC REACTION RATE k_1(SEC.$^{-1}$) $\times 10^4$	MEAN DEVIATION	RATIO AVERAGE k_1 TO HIGH PRESSURE k_1
(NO EXTRA SURFACE)							
3	40.32	0.0470	0.0423	6	2.50	0.45	0.947
1	40.04	0.0431	0.0370	7	2.48	0.13	0.973
2	40.04	0.0142	0.0131	7	2.22	0.24	0.871
5	40.32	0.00970	0.00780	5	2.31	0.25	0.875
4	40.32	0.00850	0.00500	4	1.80	0.18	0.682
6	40.32	0.00725	0.00420	4	1.60	0.10	0.606
7	40.32	0.00565	0.00443	4	1.87	0.24	0.708
(WITH TUBES IN FLASK)					RANGE OF k_1		
10	40.32	0.0401	0.0335	7	2.96 to 1.90		1.02
8	40.32	0.0186	0.0119	5	4.69 to 2.62		1.26
9	40.32	0.0111	0.00643	6	4.24 to 2.37		1.33
11	40.32	0.01005	0.00777	5	4.58 to 1.48		1.07

These results, which seem *a priori* to be of a very reasonable nature, are in complete contradiction with the results of all previous investigators. Thus Hirst and Rideal,[3] working in the range 1.450 to 0.035 mm., report an increase in specific rate with decreased pressure which became several fold at their lowest pressures. Hibben,[4] working in the range 0.18 to 0.03 mm. found a slight increase in specific rate, which he believed to be within his experimental error. Sprenger,[5] working in the range 0.05 to

0.01 mm., came to the conclusion that nitrogen pentoxide decomposes at approximately its high pressure rate when first introduced into the reaction flask, and later, with a considerable fraction of the original pentoxide still remaining, ceases to decompose at all. Finally, Rice, Urey and Wilson[6] report as a preliminary result a falling-off in the rate even at pressures as high as several centimeters.

We believe that an important factor in explaining these extraordinarily discordant results of previous investigators is to be found in the effect of wall surface which was studied by us in runs Nos. 8 to 11, where we increased the surface to volume ratio from 0.14 cm.$^{-1}$ to 0.73 cm.$^{-1}$. This increase in surface appears to have the effect at very low pressures of greatly increasing the rate especially at the beginning of the run. Since none of the previous investigators used a reaction vessel anywhere near as large as our 45-liter flask it appears to us quite possible that their results were seriously affected by the surface of the vessel. This consideration would explain the high results of Hirst and Rideal, especially as they found a large falling-off in the specific reaction rate with the progress of the run. In the case of Hibben, who found only a slight increase in rate in the region where we found some decrease, we should have to conclude on this basis that he had a sufficiently inactive surface so that the effect was not great with his surface to volume ratio, which we estimate to have been about 0.5 cm.$^{-1}$, thus lying in between the two conditions that we studied. As for the results of Sprenger, in the absence of satisfactory confirmatory experiments we cannot in the least accept his conclusion that his nitrogen pentoxide stopped decomposing with a considerable fraction of the original pentoxide still present, and believe that the discrepancy between his initial and final pressure measurements is to be accounted for on the basis of impure nitrogen pentoxide at the start. If rates for his experiments are calculated on this basis they show the same kind of behavior as those of Hirst and Rideal, and could receive the same explanation. Finally the preliminary result of Rice, Urey and Wilson, as they themselves appear to believe, may be affected by analytical inaccuracies and hardly needs discussion at the present time. Although we have discussed above the possibility of wall effect, as a factor in explaining the discordant results of other investigators, this may not be the only factor, especially as there are necessarily many unsatisfactory features about any technique that can be used in measuring the low pressure rate of decomposition of nitrogen pentoxide.

It remains to discuss some of the possible sources of error in our own results as given by runs Nos. 1 to 7 made in the large 45-liter flask in the absence of extra surface. In the first place, there must have been some residual effect of the walls even in this flask and no certain estimate of the extent of this residual effect can be given. Nevertheless, since our

surface to volume ratio was so very much smaller in these runs than in those with the tubes, and the specific rates obtained showed none of the dropping off with progress of the run characteristic of the surface effect, we are inclined to believe that our results are not greatly affected by the surface. It is of interest to note, however, that any correction which should be made for the action of the walls would be in the nature of slightly accentuating the falling-off of the specific rates below their high pressure value. In the second place, it was unfortunate that we had no runs at these very low pressures in which the calculated initial pressure of pentoxide was equal to the total initial pressure. The purity of the pentoxide, calculated after the short interval necessary for the pressure measurement, varied in runs Nos. 1 to 7 from 58 to 92%, in spite of the fact that we endeavored to get pentoxide as free as possible from its decomposition products. Nevertheless, this wide range of purities did not appear to have any effect on the consistency of our results. It is of interest to note also in this connection, however, that if we assumed that the pentoxide was really purer than calculated, the calculated specific rates would have shown even a greater falling-off below the high pressure value. For the present we do not see other possible sources of error which might be serious.

Finally, a word may be said about the bearing of these results on the theory of chemical activation by collision as developed by Rice and Ramsperger, and by Kassel. Qualitatively, the result is very satisfactory since it shows a falling-off in the rate of this unimolecular decomposition at low pressures, as is to be expected when the pressure becomes so low that the molecules cannot be activated fast enough by collisions to maintain the high pressure rate. Quantitatively, however, the result is hard to explain since if we stay within the range of kinetic theory diameters for the molecules we should expect the falling-off to occur at higher pressures. Nevertheless, the quantitative computations of the theory involve considerable complexity and possibility of choice as to assumptions. In any case, we are inclined to regard the falling-off that we have found as evidence in favor of the idea that nitrogen pentoxide receives its energy of chemical activation, at least to a considerable extent, by interaction among the molecules even if these interactions could not be called collisions in the kinetic theory sense.

[1] Ramsperger, Nordberg and Tolman, *Proc. Nat. Acad. Sci.*, **15**, 453 (1929).
[2] Olson and Hirst, *J. Am. Chem. Soc.*, **51**, 2378 (1929).
[3] Hirst and Rideal, *Proc. Roy. Soc.*, **109A**, 526 (1925).
[4] Hibben, *J. Am. Chem. Soc.*, **50**, 940 (1928).
[5] Sprenger, *Zeit. physik. Chem.*, **136**, 49 (1928).
[6] Rice, Urey and Wilson, reported by Rice at the Minneapolis meeting of the American Chemical Society, September, 1929.

A NEW EQUATION OF STATE FOR FLUIDS. IV. AN EQUATION EXPRESSING THE VOLUME AS AN EXPLICIT FUNCTION OF THE PRESSURE AND TEMPERATURE

By James A. Beattie

Research Laboratory of Physical Chemistry, Massachusetts Institute of Technology*

Communicated November 20, 1929

Equations of state have almost universally been of such a form that they can be solved explicitly for the pressure but not for the volume. This is true of the Beattie-Bridgeman equation of state[1] of which the virial form is

$$p = \frac{nRT}{V} + \frac{n^2\beta}{V^2} + \frac{n^3\gamma}{V^3} + \frac{n^4\delta}{V^4} \qquad (1)$$

$$\beta = RTB_0 - A_0 - Rc/T^2$$
$$\gamma = -RTB_0b + A_0a - RB_0c/T^2$$
$$\delta = RB_0bc/T^2$$

in which R is the ideal gas constant; A_0, a, B_0, b and c characteristic constants for each gas; and V the volume of n moles of gas. This equation has been applied[2] to the compressibility data on twelve gases, and calculations[3] of the weight of a normal liter have shown that for some thermodynamic computations the values of the constants for nitrogen and carbon dioxide may be used for carbon monoxide and nitrous oxide, respectively, a result not unexpected in view of Langmuir's[4] theory of isosteres. The equation not only represents the compressibility data over wide ranges of temperature and density about as well as two different investigators agree on the values but also reproduces well the other thermodynamic[5] properties of gases including the inversion temperature of the Joule-Thomson effect.

This equation of state has been generalized[6] to apply to gas mixtures by use of essentially the same method of calculation of the values of the parameters of the mixture from the composition and the constants of the pure component gases as that used by Lurie and Gillespie,[7] Keyes[8] and others:

$$R_n = \Sigma_1(n_1)R \qquad A_{0n} = [\Sigma_1(n_1\sqrt{A_{01}})]^2 \qquad a_n = \Sigma_1(n_1a_1)$$
$$B_{0n} = \Sigma_1(n_1B_{01}) \qquad b_n = \Sigma_1(n_1b_1) \qquad c_n = \Sigma_1(n_1c_1). \qquad (2)$$

In these relations the subscript 1 refers to the values of the constants for Substance 1, the subscript n to the values for the mixture whose composition is expressed in terms of the mole numbers n_1, n_2, ..., and the sum-

mation extends over all of the gases in the mixture. The equation of state of a gas mixture is obtained by substituting the parameters for the mixture into equation (1), giving the relation:

$$p = \frac{R_n T}{V} + \frac{\beta_n}{V^2} + \frac{\gamma_n}{V^3} + \frac{\delta_n}{V^4} \tag{3}$$

$$\beta_n = R_n T B_{0n} - A_{0n} - R_n c_n / T^2$$
$$\gamma_n = -R_n T B_{0n} b_n + A_{0n} a_n - R_n B_{0n} c_n / T^2$$
$$\delta_n = R_n B_{0n} b_n c_n / T^2$$

in which V is the volume of $\Sigma_1(n_1)$ moles of the mixture. Equation (3) represents well[6] the data of Keyes and Burks[8] on the compressibility of mixtures of methane and nitrogen; and Gillespie[9] has shown that the expansion of gases on mixing at low pressures can be calculated from equation (3) for binary mixtures of methane-nitrogen, hydrogen-nitrogen, ethylene-argon and ethylene-oxygen. In general it may be said that (3) represents the properties of a gas mixture about as well as (1) reproduces the properties of pure gases.

General relations for the thermodynamic properties of real gases and gas mixtures in terms of the independent variables V and T have been derived[10] and these were integrated by use of equations (1) and (3). General relations for the thermodynamic properties of real gases and gas mixtures in terms of the independent variables p and T have been derived by Gillespie[11] and by the author.[10] Lurie and Gillespie[7] integrated these relations by graphical methods. It would be convenient to have available an equation of state of the form $V = f(p,T)$ by means of which the integration of these relations could be carried out, expressing the thermodynamic properties in terms of elementary functions of p and T.

Kinetic theory considerations always lead to an equation of the form $p = f(V,T)$ such as equation (1). In order to obtain an equation of the type $V = f(p,T)$, we can partially solve (1) for the volume and replace n/V by p/RT in the higher order terms. There results for the virial form:

$$V = \frac{nRT}{p} + \frac{n\beta}{RT} + \frac{n\gamma}{R^2T^2}\,p + \frac{n\delta}{R^3T^3}\,p^2 \tag{4}$$

where β, γ and δ are the functions of the equation of state constants and the temperature given in equation (1). The form of this equation which is most suited for a general discussion is

$$V = n[\pi + B](1 - \epsilon) - \frac{nA}{RT} \tag{5}$$

$$A = A_0(1 - a/\pi)$$
$$B = B_0(1 - b/\pi)$$
$$\epsilon = c/\pi T^3$$

where $\pi = RT/p$ and has the dimensions of molal volume.

In order to obtain the best representation of the compressibility measurements by use of equation (4) or (5), the constants A_0, a, B_0, b and c should be redetermined from the data, and they would have somewhat different values from those which satisfy the data when substituted into equation (1). A method for evaluating the constants of (4) and (5) from the experimental measurements and one which was successfully applied[2] for the constants of equation (1) is evident when (5) is written in the form:

$$\frac{V}{n} + \frac{c\psi}{\pi T^3} = \psi - \frac{\varphi}{RT} \qquad (6)$$

$$\psi = \pi + B_0 - B_0 b/\pi$$
$$\phi = A_0 - A_0 a/\pi$$

By use of methods of smoothing described in reference (2), values of the molal volume at each measured isotherm can be determined from the data for a series of values of π, the same values of π being used for each isotherm. The temperatures and volumes along any one constant π curve should give a straight line when the function $\left(V + \frac{c\psi}{\pi T^3}\right)$ is formed and plotted against $1/T$. The value of c which leads to a straight line must be determined by trial and the method depends on a knowledge of the intercept ψ of the constant π line, so that several approximations are necessary, but as in the case of the determination of c for equation (1) the convergence is rapid. The other constants can be determined from the intercepts ψ and slopes $-\phi/R$ of the constant π lines. When $(\psi - \pi)$ is plotted against $1/\pi$, a straight line should result, the intercept being B_0 and the slope $-B_0 b$; and when ϕ is plotted against $1/\pi$ another straight line with an intercept A_0 and a slope $-A_0 a$ should result. A preliminary study of the data indicates that constants which give a good representation of the experimental measurements can be determined in this manner.

It is of interest to find out how well the values of the constants which have previously been determined for use in equation (1) will reproduce the experimental data when used in (4). In table 1 the virial parameters calculated[3] from the constants of (1) are given for 0°C. for the eleven substances which are gaseous under one atmosphere pressure at this temperature. These values were then used in equation (4) to calculate the volumes. The results are given in table 2. In the second column of this table are given the critical temperatures of the gases. In the

succeeding twelve columns are given the percentage deviation of the calculated from the observed densities. In these calculations, use was made of the smoothed values of the pressures which had already been determined;[2] for He, Ne, A, H_2, N_2, O_2, and air the smoothed data of Holborn and Otto[12] being used, and for CO_2, CH_4, C_2H_4 and NH_3, those of Andrews,[13] Keyes and Burks,[14] Amagat[15] and Meyers and Jessup,[16] respectively. The last three columns of the table give the average percentage error in the calculation of the density by equation (4), the average percentage deviation in the calculation of the pressure by (1) the same values of the virial parameters being used in each case, and the maximum pressure at which the comparison was made in each case.

TABLE 1

VIRIAL PARAMETERS AT 0°C.

The values listed were calculated from the constants given in references (2), and were determined for use in equation (1).

Units: atmospheres, liters per mole, degrees Kelvin (T°K. = t°C. + 273.13). For all gases RT = 22.41305.

	β	$10^3\gamma$	$10^6\delta$
Helium	0.29214	1.292	0.0
Neon	0.24810	4.644	0.0
Argon	−0.47553	27.457	0.0
Hydrogen	0.27172	19.467	−0.5
Nitrogen	−0.25974	40.669	−16.1
Oxygen	−0.50752	31.399	10.3
Air	−0.31547	34.303	−24.2
Carbon dioxide	−3.38451	111.131	5502.6
Methane	−1.16581	54.224	−125.1
Ethylene	−3.67695	177.057	1090.9
Ammonia	−6.87316	82.131	34236.5

For all gases which are above their critical temperature, equation (4) gives a good representation of the volume. For gases below their critical temperature and especially in the region which is near the critical temperature and also the saturation line, the agreement is not good, though considerably better than the ideal gas law, which gives errors of 28%, 23% and 48%, respectively, for the highest densities listed for CO_2, CH_4 and C_2H_4.

Equation (4) can be generalized to apply to gas mixtures, the resulting relation being:

$$V = \frac{R_n T}{p} + \frac{\beta_n}{R_n T} + \frac{\gamma_n}{R_n^2 T^2} p + \frac{\delta_n}{R_n^3 T^3} p^2, \qquad (7)$$

where β_n, γ_n and δ_n are the functions of the composition, temperature and the equation of state constants of the pure component gases given in equation (3). This relation will evidently represent the data on gas mixtures as well as (4) reproduces the measurements on the pure gases.

TABLE 2

DEVIATIONS OF THE DENSITIES CALCULATED BY EQUATION (4) FROM THE OBSERVED VALUES AT 0°C.

The values of the virial parameters used are given in table 1.

The pressures corresponding to the densities listed are given in references (2).

GAS	CRITICAL TEMP. °C.	DENSITY IN MOLES PER LITER: 0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	AV. DEV. EQ. 4	AV. DEV. EQ. 1	MAXIMUM PRESSURE
		OBSERVED MINUS CALCULATED DENSITY IN PER CENT														
He	−268	0.0	0.1	0.2	0.3	0.3	0.4	0.5	0.6					0.30%	0.21%	94.1
Ne	−229	0.0	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3				0.18	0.07	106.3
A	−122	0.0	0.1	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.5			0.25	0.04	103.5
H_2	−240	0.0	−0.1	−0.1	−0.1	0.0	0.2	0.3	0.6					0.18	0.09	95.2
N_2	−147	0.0	−0.1	−0.1	−0.2	−0.2	−0.1	−0.1	−0.1	−0.1				0.11	0.11	99.3
O_2	−119	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.3	0.4	0.5			0.17	0.08	103.2
Air	−141	0.0	−0.1	−0.1	−0.1	−0.1	−0.2	−0.2	−0.2	−0.2				0.13	0.16	97.8
CO_2	31	0.6	2.3	5.3	9.6									4.5	0.12	32.2
CH_4	−83	0.0	0.3	0.6	1.0	1.4	1.9	2.3	2.8	3.2	3.6	3.9	4.3	2.1	0.09	104.2
C_2H_4	10	0.6	2.8	6.1	10.7	16.2	22.5	29.5						12.6	0.58	40.4
NH_3	132	0.2[a]	0.5[b]											0.35	0.23	2.5

[a] For $d = 0.045$.

[b] For $d = 0.117$.

The general equation for the thermodynamic properties of real gases and their mixtures which express these properties as functions of the independent variables p and T can now be integrated, giving relations for pure gases in terms of p, T, and their equation of state and thermal constants; and relations for gas mixtures in terms p, T, the composition and the equation of state and thermal constants of the pure component gases.

Thus by use of a single set of equations of state constants, together with thermal data on pure gases, it is possible to give equations which lead to a numerical solution for the thermodynamic properties of real gases and their mixtures in terms of either set of independent variables V, T or p, T. Some equations have a simpler form in terms of one set of variables and others in the other set. The experimental data, however are usually given in terms of the independent variables p and T.

* Contribution No. 230.

[1] Beattie, J. A., and Bridgeman, O. C., *J. Amer. Chem. Soc.*, **49**, 1665 (1927); *ibid.*, **50**, 3133 (1928).

[2] Beattie, J. A., and Bridgeman, O. C., *Proc. Amer. Acad. Arts Sci.*, **63**, 229 (1928) [helium, neon, argon, hydrogen, nitrogen, oxygen, air, carbon dioxide, methane, ethyl ether]; Gillespie, L. J., *J. Phys. Chem.*, **33**, 354 (1929) [ethylene]; Beattie, J. A., and Lawrence, C. K., *J. Amer. Chem. Soc.*, **52**, Jan. (1930) [ammonia]. A theoretical basis for the equation is given in the first reference.

[3] Beattie, J. A., and Bridgeman, O. C., *J. Amer. Chem. Soc.*, **50**, 3151 (1928).

[4] Langmuir, I., *J. Amer. Chem. Soc.*, **41**, 868, 1543 (1919).

[5] Bridgeman, O. C., *Physical Review*, **34**, 527 (1929) [Joule-Thomson effect and heat capacity at constant pressure of air]; Beattie, J. A., *Jour. Math. Physics*, (in press) [Joule-Thomson effect and heat capacity at constant pressure of ammonia].

[6] Beattie, J. A., *J. Amer. Chem. Soc.*, **51**, 19 (1929).

[7] Lurie, E., and Gillespie, L. J., *J. Amer. Chem. Soc.*, **49**, 1146 (1927).

[8] Keyes, F. G., and Burks, H. G., *J. Amer. Chem. Soc.*, **50**, 1110 (1928).

[9] Gillespie, L. J., *Physical Review*, **34**, 352 (1929).

[10] Beattie, J. A., *Physical Review*, **31**, 680 (1928); *ibid.*, **32**, 691, 699 (1928).

[11] Gillespie, L. J., *J. Amer. Chem. Soc.*, **47**, 305 (1925); *ibid.*, **48**, 28 (1926).

[12] Holborn, L., and Otto, J., *Zeit. Physik*, **33**, 1 (1925).

[13] Andrews, T., *Phil. Trans.*, **159**, 575 (1869); *ibid.*, **166**, 421 (1876).

[14] Keyes, F. G., and Burks, H. G., *J. Amer. Chem. Soc.*, **49**, 1403 (1927).

[15] Amagat, E. H., *Ann. Chim. Phys.*, (6) **29**, 68 (1893).

[16] Meyers, C. H., and Jessup, R. S., *Refrigerating Engineering*, **11**, 345 (1925).

CRYSTAL STRUCTURES OF SOME DERIVATIVES OF DIPHENYL

By George L. Clark and Lucy W. Pickett

Department of Chemistry, University of Illinois

Communicated December 5, 1929

A two-fold interest attaches itself to x-ray studies of structure of diphenyl derivatives in that they offer a mode of attack on the problem of the shape and symmetry of the benzene ring, and also on that of the stereoisomerism of 2,2′-6,6′ substituted diphenyls now being intensively investigated by Roger Adams and associates. An x-ray study of a series of derivatives of diphenyl is being carried out with these interests, and unique solutions have been reached for diphenyl, dimesityl and *o*-tolidine.

The analysis in each case has been carried out with the use of single crystals of one or two millimeters for the longest dimension. Although sharper diffraction spots are obtained from smaller crystals, the difficulties of mounting and the length of time of exposure increase greatly, so that this size was found to give the most satisfactory results. Laue photographs were taken using the unfiltered radiation from a molybdenum target tube; for the rotation and oscillation patterns the characteristic K radiation of copper was employed. No filter was used inasmuch as the α-radiation could be easily distinguished from the β by the great difference in intensity, and served as a check on calculations. The crystallographic system was determined in each case with the x-ray goniometer as described by Clark and Yohe,[1] with further verification from Laue patterns and optical examinations. The dimensions of the unit cell were obtained from rotation photographs about each of the three axes of the crystal, while identification of the reflections corresponding to each set of planes could be made with certainty from sets of oscillating photographs, each covering a range of 15°. This is perhaps the first time that these oscillation patterns have been used to such good advantage. Typical photographs of each type are reproduced in this note.

Diphenyl.—Although an analysis of diphenyl was recently published by Mark,[2] the following results, obtained independently, are presented in confirmation of his conclusions. The crystal unit was found to be monoclinic with the dimensions $a = 8.11$, $b = 5.67$, $c = 9.57$ and $\beta = 94.5°$. From the facts that 010 is halved and also $h0l$ where h is odd, the crystal is assigned to the monoclinic prismatic class with the space group C_{2h}^5 based on the simple monoclinic lattice.[3] Inasmuch as there are two molecules in the unit cell, conditions of symmetry demand that the molecule should have a center of symmetry only and hence a staggered ring is probable. These results are identical with those of Mark except for slight differences in the length of the axes.

Dimesityl.—Dimesityl crystallizes from an alcohol water mixture in thin rectangular plates or in needles. Rotation and oscillation photographs taken about each of the three axes indicate that the unit cell is monoclinic prismatic with $\beta = 96°30'$; $a = 8.21$, $b = 8.58$, $c = 22.25$. The space group, based on the simple monoclinic lattice, Γ_m, is C_{2h}^5, and the unit cell contains four asymmetric molecules. The following data are presented in support of these conclusions:

Rotation about a axis

n	Distance of line from equator		a
1	0.95 cm.		8.238
2	2.03		8.180
3	3.40		8.211
		Ave.	8.21

Rotation about b axis

			b
1	0.915		8.545
1	0.815		8.638
2	1.915		8.604
2	1.72		8.538
		Ave.	8.58

Rotation about c axis

			c
1	0.345		22.20
2	0.70		22.20
2	0.63		22.26
3	1.06		22.34
		Ave.	22.25

The density of dimesityl, determined by immersing a crystal in an aqueous salt solution of the same density, is 1.020. Thence

$$\sin 83'30' \times 10^{-24} \times 8.21 \times 8.58 \times 22.25 = \frac{238.176 \times n}{6.06 \times 10^{23} \times 1.02}$$

$n = 4.034$ or 4 molecules in the unit cell.

Oscillation photographs about the 001 axis show that the 100 spacing is present in all orders, but that the 010 is halved. Photographs about the 010 axis show that the 001 spacing is halved, while a photograph about 001 shows that 101, 201, 103, 203, etc., are missing. All other reflections seem to be present. Thus it is obvious that the planes $h0l$ where l is odd are halved and 010 is halved. The only space group which meets this condition is C_{2h}^5. Since there are four molecules in the unit cell, the molecule must be asymmetric.

It is of interest that the space group is the same as that of diphenyl. In the latter case, however, there are but two molecules in the unit cell

and each molecule has a center of symmetry. The introduction of the three methyl groups in each ring, though in seemingly symmetrical positions evidently destroys the symmetry of the molecule.

The equivalent points in the space group C_{2h}^5 are represented in figure 1 which is a diagram of the unit cell. The positions marked 1 at the eight corners of the cell are equivalent; those marked 2 are separated from the former by a rotation of 180° and a translation of $^1/_2$; those marked 3, at

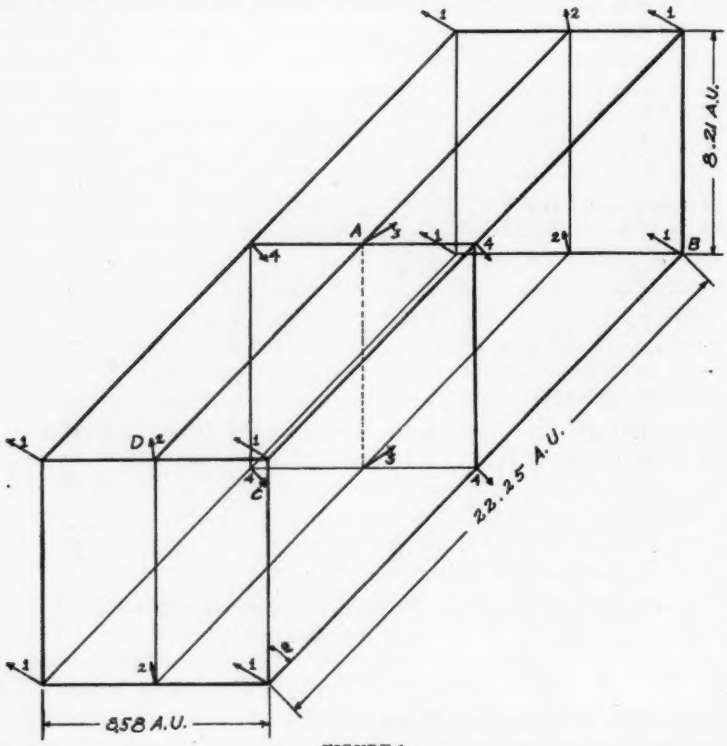


FIGURE 1

top and bottom have a center of symmetry with the elements at 1 and are related to those at 2 by reflection in a glide plane; those at 4 have a center of symmetry with the ones at 2, and are enantiomorphous to those at 1. The coördinates of these points may be expressed:[4] x, y, z; $\bar{x}, \bar{y}, z + {}^1/_2$; $x + {}^1/_2, y, \bar{z}$; ${}^1/_2 - x, \bar{y}, {}^1/_2 - z$.

The exact position of the molecules is not known from the present data. It is probable that in the case of diphenyl, the two molecules occupy positions with their axes parallel to AB and CD, respectively, in which case

the length of the axis would be about 9.8. For dimesityl, however, the long axis of the molecule could hardly be in this direction, since two molecules would have to have a combined length of 14.46.

3,3′-Diamino-dimesityl.—Upon the theory that certain substituent groups on the two phenyl groups are sufficiently large in their spacial spheres of influence so that free rotation around the diphenyl linkage would be prevented by interference at 2 and 2′, and 6 and 6′ positions dimesityl with interfering CH_3 groups should come under this classification. Two optical isomers of course could not be expected on account of the

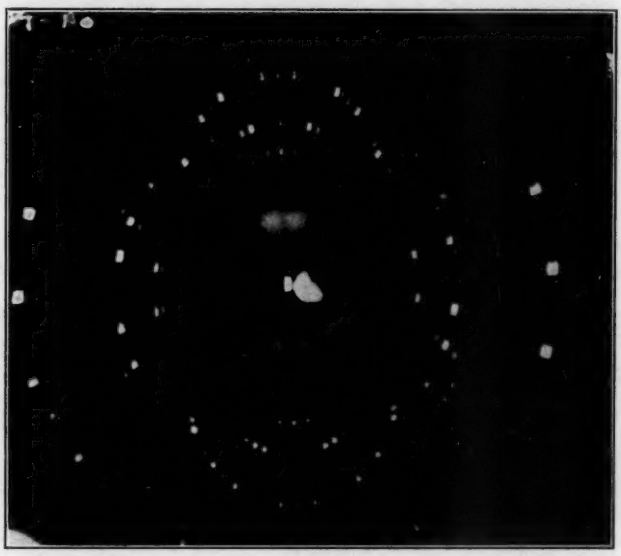


FIGURE 2

symmetrical placing of groups all alike (six methyl groups). By introducing an additional single substituent group on each phenyl, thus unsymmetrically loading the molecule, the interference of 2,2′-6,6′ methyl groups should be manifest in the resolution of optical isomers.[5] Such a compound is 3-3′ diamino-dimesityl which may be resolved into *d* and *l* forms. X-ray patterns prove that these *d* and *l* forms are identical, and that the crystalline structure is similar to that of dimesityl. A detailed analysis will be presented later.

o-Tolidine.—Ortho-tolidine, or 4,4′ diamino- 3,3′ dimethyldiphenyl crystallizes from 95% alcohol in rectangular plates slightly brownish in color. Examination of single crystals by Laue, rotation and oscillation

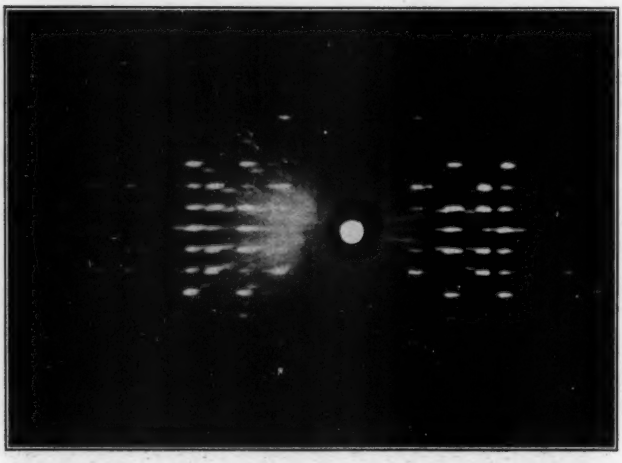


FIGURE 3

photographs showed that the unit cell is orthorhombic, contains four molecules, and has the dimensions $a = 6.50$, $b = 7.48$ and $c = 23.62$ A.U.

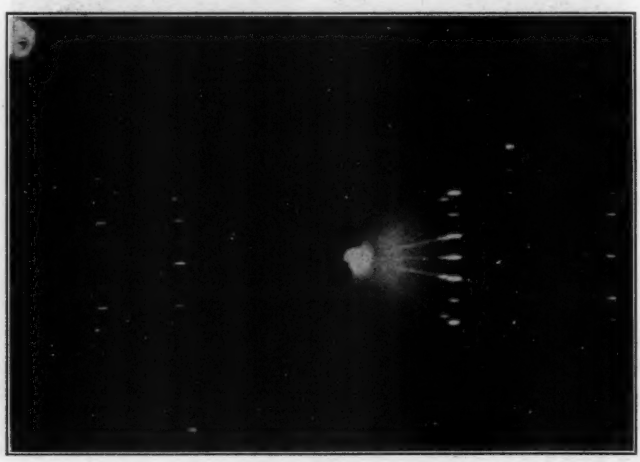


FIGURE 4

Figure 2 represents a Laue photograph of *o*-tolidine with the beam of x-rays parallel to the *a* axis; the planes with indices from 1,6,0 to 1,0,20 are to be identified on this photograph. A typical rotation diagram is illustrated in figure 3 where the *c* axis is the axis of rotation and its length

may be calculated from the distance between the layer lines on which the spots lie. Figure 4 is one of a series of oscillation photographs covering a range of 25° to 40° taken about the *c* axis of tolidine. This method proved the most valuable in the identification of reflections, inasmuch as in this way the most complicated rotation diagram could be analyzed into a number of simpler diagrams on which spots could be identified with certainty because the position of the crystal was known and spots are not likely to be superimposed. Thus in figure 4, the vertical line of spots at the right of the central beam refer to the planes 110, 111, 112, 113, etc. Those at the left near the central beam are 210, 211, 212, 213, 214, interferences, etc., and the line of spots further removed corresponds to 310, 311, 312, 313, 314, etc., planes.

The dimensions of the unit cell were obtained from the following measurements of rotation photographs.

Rotation about *a* axis

n	Distance of line from 0 line	*a*
1	1.215 cm.	6.52
2	2.695	6.49
2	2.70	6.47
2	2.35	6.53
	Ave.	6.50

Rotation about *b* axis

		b
1	1.05	7.48
2	2.24	7.51
2	2.26	7.45
2	2.00	7.48
	Ave.	7.48

Rotation about *c* axis

		c
1	0.325	23.68
2	0.65	23.86
3	1.00	23.50
4	1.36	23.59
5	1.725	23.56
6	2.125	23.57
	Ave.	23.62

The density of the crystal was experimentally determined as 1.215. From these values the number of molecules in the unit cell may be calculated:

$$6.50 \times 7.48 \times 23.62 \times 10^{-24} = \frac{212.2 \times n}{6.06 \times 10^{-23} \times 1.215}$$

$$n = 3.985, \text{ or } 4.$$

The $00l$ planes where l is odd and the $h00$ planes where h is odd are definitely absent, while the $0k0$ planes where k is odd are so weak that they are only detectable on an overexposed photograph. No other group is consistently absent. Hence the space group is probably Q^4, also symbolized as V^4; if the $0k0$ planes are present and weakened by some other cause than that of space group considerations, the group Q^3 in the alternative. Both groups are based on a simple orthorhombic lattice, Γ_0, and require four unsymmetrical molecules in the unit cell.

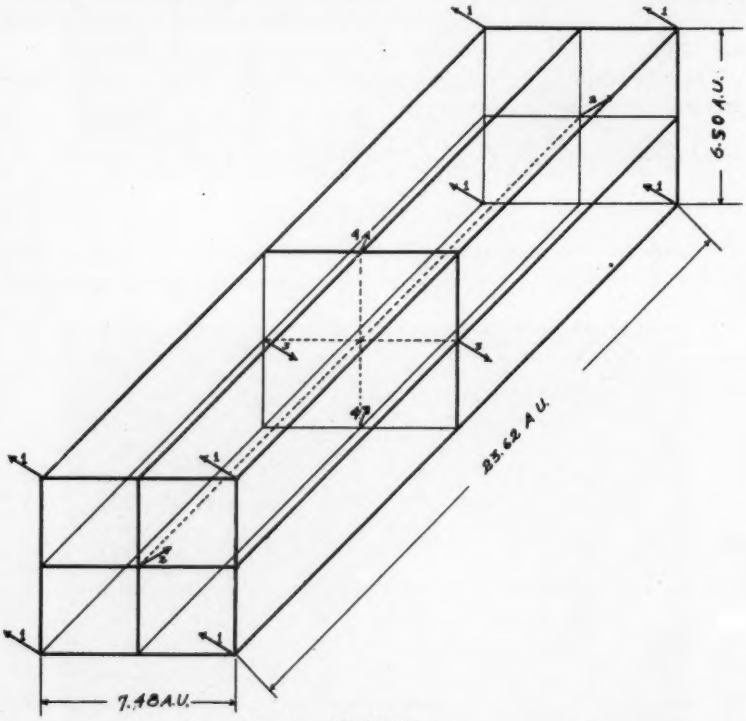


FIGURE 5

The four equivalent positions for the space group V^4 may be designated by the coördinates xyz; $x + 1/2, 1/2 - y, \bar{z}$; $\bar{x}, y + 1/2, 1/2 - z$; $1/2 - x, \bar{y}, z + 1/2$.

Figure 5 represents the unit cell with equivalent points at the corners and three other sets at the centers of opposite faces. Those at 1 and 3, and at 2 and 4 are related by dyad screw axes parallel to the edge with

spacing 23.62; those at 1 and 2, and at 3 and 4 have dyad screw axes parallel to the edge with spacing 6.50.

Diphenic Acid.—The analysis of this compound is not complete. The results obtained show that the crystal is orthorhombic with the dimensions $a = 14.12$, $b = 11.90$, $c = 13.75$, and that there are eight molecules in the unit cell. This indicates an arrangement quite different from the other compounds studied as the unit cell is approximately half as long in one dimension and twice as wide in the other two. It is possible that this is caused by the attraction and consequent doubling of COOH groups such as is found in the case of aliphatic compounds.

Other compounds now being investigated are: difluoro-diphenic acid, hexachloro-diphenyl, octamethyl diphenyl, diphenyl benzene. Thanks are due Professor Roger Adams for suggesting this research and for providing the necessary materials.

[1] *J. Am. Chem. Soc.*, **51**, 2796 (1929).

[2] *Zeit. Krist.*, **70**, 285 (1929).

[3] Astbury and Yardley, *Phil. Trans. Roy. Soc. London*, **224A**, 221 (1924).

[4] Wyckoff, *Analytical Expression of the Results of the Theory of Space Group*, Carnegie Institute, Washington, p. 51.

[5] See papers by Roger Adams and associates on diphenyl derivatives in *J. Am. Chem. Soc.*

THE TEMPERATURE COEFFICIENT OF THE PHOTOCHEMICAL FORMATION OF HYDROGEN CHLORIDE

BY OLIVER R. WULF

BUREAU OF CHEMISTRY AND SOILS, WASHINGTON, D. C.

Communicated November 19, 1929

In the theory of the temperature coefficient of photochemical reaction developed by Tolman[1] the probability that a molecule will react, once it has acquired by absorption the energy necessary for reaction, is taken to be not markedly dependent on the temperature. In the case of unimolecular reactions it is important to notice that one then has simply the theory of the temperature coefficient of light absorption. Kistiakowsky[2] has discussed this in his recent book. It is especially important to remember that this remaining temperature coefficient of light absorption can appear experimentally to its full extent only in those cases of small fractional reduction of the intensity of the incident beam. The temperature coefficient of light absorption for molecules is a quantity regarding which the theory and the data of band spectra should be able to give complete information. It will be noted that in Tolman's theory extended

to bimolecular reactions a new possible source of temperature coefficient enters in the thermal activation of the second reactant, but that the probability of reaction, once the necessary energy has been acquired and collision occurs, is still left as being probably not markedly dependent on the temperature. Where there is no necessity of activating the second reactant the theory again becomes essentially the theory of the temperature coefficient of light absorption.

It is the purpose of this paper to calculate the variation of photochemical temperature coefficient with wave-length under these simplified conditions using a reaction which, over certain regions, seems to be illustrative.

The reaction is the photochemical combination of hydrogen and chlorine. It is clear that chlorine is the only photochemically active constituent. It is, furthermore, very probable that the elementary photochemical act is the dissociation of the chlorine molecule. Rollefson[3] has recently studied and discussed the status of this reaction, and has developed the probable mechanism. It will be assumed in what follows that the necessary primary photochemical act is the dissociation of the chlorine molecule.

Padoa and Butironi[4] investigated the temperature coefficient of this reaction in green, blue, violet and ultra-violet light at 10°, 20°, 30° and 40°. Filters were used to isolate these regions, from a continuous source in the first three cases, and from a mercury arc in the ultra-violet. The authors give the spectral range transmitted by each, and the temperature coefficient, expressed as the ratio of the rates separated by an interval of ten degrees, for the photochemical reaction in each of these regions. Their results are summarized in table 1.

TABLE 1

	REGION	TEMP. COEFF. 10°
Green	5500–5300 Å	1.50
Blue	4900–4700	1.31
Violet	4600–4400	1.21
Ultra-violet	4000–3500	1.17

From the weakness of chlorine's absorption in the visible, it is very probable that for the first three cases the fractional reduction of intensity of the incident beam was small; surely so in the first two cases.

The band spectrum of chlorine in the visible and near ultra-violet is well known from the work of Kuhn[5] and others. Absorption from at least the first five vibrational levels of the normal molecule is observable. One can say from which particular vibrational levels the absorption of chlorine in the above regions at ordinary temperatures originates, and the energy of these levels is known. This is sufficient to determine the temperature coefficient of such absorption. Indeed it is partly by a process the reverse of this that the allocation of absorption to the various vibrational levels is accomplished. And so from the positions of the four

spectral regions investigated by Padoa and Butironi, we should be able to say what levels in chlorine are responsible for the photochemically active absorption, and to determine from the band spectra data whether indeed this absorption has a temperature coefficient corresponding to that found by these investigators.

In table 2 the energies of vibration in calories for the first six vibrational levels of the normal chlorine molecule are given, along with the beginning of the continuous spectrum at the limit of the band progression built upon each of these levels, that is, the point at and beyond which, toward shorter wave-lengths, the molecule absorbing in this level dissociates in the elementary act of light absorption. In the fourth column is given approximately the fractions of the molecules that exist in these levels at 298°K., and in the last column the temperature coefficient of light absorption by each level. The energy of the lowest vibrational state has been taken as one-half unit, as is indicated by the new mechanics. The temperature coefficient of the concentration of molecules in the respective levels computed using the Boltzmann distribution is, however, the same as for the levels according to the old mechanics. The calculations have been carried so that the values given for the temperature coefficient are correct within about ± 1 in the last figure.

TABLE 2

QUANTUM NUMBER n''	ENERGY, CALORIES PER MOL.	BEGINNING OF CONTINUOUS SPECTRUM, Å.	BOLTZMANN QUOTA, 298°K.	TEMP. COEFF. (FOR 10°)
0	813	4659	0.934	0.994
1	2420	4785	6.2×10^{-2}	1.09
2	4002	4916	4.3×10^{-3}	1.19
3	5556	5052	3.1×10^{-4}	1.31
4	7085	5193	2.3×10^{-5}	1.42
5	8586	5339	1.9×10^{-6}	1.55
6	10064	5492	1.5×10^{-7}	1.68

While we shall be anxious to know the extent to which the observations agree with the theory for all four regions, perhaps the greatest interest centers around the longest wave-lengths used, since here the temperature coefficient is the greatest and the chlorine molecule is receiving the smallest energy. Here, if ever, additional thermal activation, for instance of the hydrogen molecule, would be necessary.

From a comparison of the tables it is evident that the only levels leading to dissociation in the region of green wave-lengths, given by Padoa and Butironi as transmitted by their filter, are the levels $n'' = 5$ and higher. Absorption from level $n'' = 6$ will be much weaker than from $n'' = 5$ in this region, not only because of the decrease in the Boltzmann quota in the higher levels, but also because of the shift to the red of the maximum of absorption for the higher levels in accordance with the Franck-Condon

diagram for a molecule which on electronic excitation undergoes a great weakening in binding. This is especially well illustrated in chlorine. The temperature coefficient, 1.55, for the absorption by level $n'' = 5$ agrees surprisingly well with that found for the photochemical reaction, 1.50, by Padoa and Butironi. It is probable that there is some tolerance in the wave-length limits transmitted by the filters since the cut-offs are never quite sharp. The chief absorption in the filtered region is in the $n'' = 4$ level, but in the discrete states, not in the continuous. The computed temperature coefficient is probably in better agreement with the photochemical observations than could reasonably be expected.[6]

For the region 4900–4700 Å (greenish-blue), referring again to the tables, the levels $n'' = 2$ and $n'' = 1$ are chiefly responsible for the active absorption with smaller contributions from the higher levels, all of whose continuous spectra overlie the wave-length region, but for which the absorption is small for the two reasons given above. Here again the maximum absorption lies in the discrete levels of $n'' = 1$, but is inactive if dissociation is necessary. It seems, therefore, that the temperature coefficient of the absorption might approach 1.2. The observed 1.31 seems higher than could probably be accounted for by absorption due to the higher levels, yet this is in the right direction. Furthermore, no active absorption can occur in this region from $n'' = 0$, in other words, all the active absorption lies in levels possessing some temperature coefficient. Probably it is fair to conclude, allowing for the possible error of measurement, that the agreement is not bad, and that certainly a good portion of the increase of the photochemical action with temperature is accounted for by the increase in absorption.

For the "violet" region (4600–4400 Å), better called blue, the active absorption could come from all levels. However, from Kuhn's observations it is known that the intensity of absorption from $n'' = 0$ is small in this region. None of the bands due to $n'' = 0$ appear in chlorine absorption and the maximum for this level lies far to short wave-lengths in the vicinity of 3400 Å. Thus here also the absorption is largely in levels possessing a temperature coefficient. It would seem that $n'' = 1$ would be much the most active. Here, then, the observed value of 1.21 seems somewhat high. It is not easy to estimate the contributions by the higher levels which would tend to raise the temperature coefficient. The change of the temperature coefficient with wave-length is of the right sign and of about the right amount. However, in this region the temperature coefficient is probably due in part to the chain character of the reaction. To a lesser extent this may of course be true of the preceding regions also.

Passing now to the ultra-violet, where apparently in Padoa and Butironi's work the group of lines in mercury at 3660 Å were the chief source, we are in a region in which the absorption is probably nearly complete

and therefore where no appreciable temperature coefficient of absorption would appear. If it were not complete, some temperature coefficient might still be expected, although the greater part of the absorption is in the lowest level, due to the overlapping of the continuous spectra of the higher n'' levels with that of the lowest. This is illustrated by the absorption measurements in bromine by Ribaud.[7] However, in the hydrogen-chlorine reaction we are now in a region where much energy is absorbed per unit volume, where much reaction occurs per unit volume, and where much energy is liberated per unit volume, due to the absorption and the exothermic character of the reaction. In this region the temperature coefficient will undoubtedly be almost wholly that of the chain mechanism. The same conclusions would hold for white light where the greater part of the absorbed light lies in this general region and is likewise taken up by the lowest vibrational level of chlorine. It is a region of wave-lengths in which the temperature coefficient of absorption is small or negligible and where the appearance of temperature coefficient in the photochemical change is not surprising because of its chain character.

Thus the photochemical combination of hydrogen and chlorine over the longer of the wave-lengths to which it is sensitive appears to illustrate Tolman's theory in its simplest form, the influence of temperature being in an important part its effect on the absorption of light by the photochemically active constituent.

For cases of small fractional absorption, elementary band spectra theory gives considerable information regarding the relation of wave-length to photochemical temperature coefficient. Most of the simple gas molecules that enter into known photochemical reactions are of the type in which the molecular binding is greatly weakened (large increase in the moment of inertia) in the electronic excitation. From the Franck-Condon diagram for such molecules, now generally familiar, it is evident that the absorption of the longer wave-lengths originates from the higher vibrational states of the normal molecule, though in some cases it may take fairly high temperatures to bring much of this out. The behavior of the halogens and the Schumann system of oxygen is illustrative. This, with its effect on the photochemical temperature coefficient, has been illustrated above in chlorine where in passing from the ultra-violet to the green-yellow, the principal absorption passes from the zeroth to the fifth vibrational state.[8]

Absorption from states higher than the lowest possesses in such molecules two maxima of intensity. With regard to the absorption toward short wave-lengths due to the higher states the measurements of Ribaud[7] in bromine show that in this case, and hence probably for all similar cases, the maxima of the continuous spectra due to the higher states overlap very closely the maximum due to the lowest, being only slightly displaced consecutively to the violet. This would indeed be anticipated. To

shorter wave-lengths of the point where a temperature increase had very little effect (actually a small decrease in absorption, as would be expected) Ribaud found that the absorption began again to increase with temperature. This evidently is the continuous absorption due to the higher vibrational states. For photochemical reactions the important consequence is that in such molecules one must expect the portion of the temperature coefficient due to light absorption to decrease with decreasing wave-length from the first appreciable absorption, but to pass through a minimum slightly less than unity, and then again to increase sharply to the disappearance of absorption.

[1] Tolman, *Statistical Mechanics with Applications to Physics and Chemistry*, Chemical Catalog Co., 1927.

[2] Kistiakowsky, *Photochemical Processes*, Chemical Catalog Co., 1928.

[3] Rollefson, *J. Am. Chem. Soc.*, **51**, 770 (1929).

[4] Padoa and Butironi, *Atti. accad. Lincei*, **25**, II, 215 (1916).

[5] Kuhn, *Z. f. Physik.*, **39**, 77 (1926); Birge, "International Critical Tables," Vol. V, p. 409 (1929).

[6] For reactions in which the rate is proportional to other than the first power of the absorbed active light, e.g., the one half power, the treatment would be modified in an evident manner.

[7] Ribaud, *Annales de physique*, **12**, 107 (1919).

[8] The above also shows that even in the cases where a well-defined initial act, such as dissociation, is necessary, we may not expect to find a sharp photochemical threshold. (Compare reference (2), page 249.) And for other than monochromatic radiation the restriction mentioned at the outset must evidently be that a photochemical temperature coefficient due to the influence of temperature on light absorption can appear to its full extent only in cases where there is small fractional reduction in the intensity of the *active* radiation.

OZONE IN LUMINESCENCE

By E. L. Nichols and Frances G. Wick

Department of Physics, Cornell University

Read before the Academy, Tuesday, November 19, 1929

That x-rays render various substances strongly thermo-luminescent has already been shown.[1] The following experiments have been made to determine the subsequent effect of ozone upon such materials.

Procedure.—A morsel of the substance to be tested was placed in a glass tube and a slow current of oxygen which had been rendered active (i.e., partly converted into O_3 by passage through an ozonizing tube) was driven over the sample for several minutes. The arrangement of the apparatus is given in figure 1. *B*, the ozonizer consists of a metal tube water cooled, within a glass tube having a broad external collar of tinfoil,

F. When a flow of oxygen was allowed to enter the apparatus at *O*, the silent discharge from an induction coil between *F* and the outer wall of the metal tube *M*,*M* produced an abundant supply of ozone. The gas, thus activated, passed over the specimen in the tube *P* and discharged

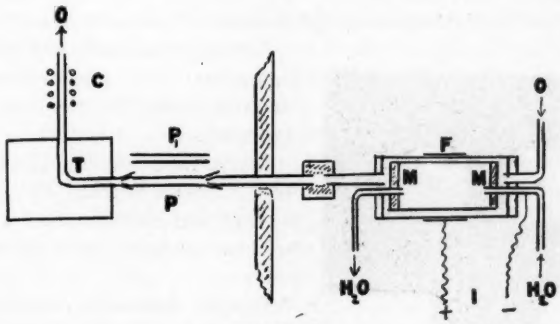


FIGURE 1
Apparatus for exposing thermo-luminescent materials to ozone.

at *O'*. The heater *C* served to reconvert the gas into ordinary oxygen. The tube connecting the ozonizer with *P* passed through a brick wall into a completely dark room, thus protecting the eyes of the observer and shielding the substance under observation from the ultra-violet rays of the electric discharge. The tube *P* containing the substance to be tested was 18 cm. long and had a bore of about 0.8 cm. It was fitted with conical ground glass joints at the ends. (For convenience of operation several such tubes, readily interchangeable, were provided.) The discharge tube *T* moved freely on slides in a direction parallel to the axis of *P*, so that the latter could be easily mounted or removed. A control sample for comparison was in the dark room and connected with *O*, so that the flow of gas could be conveniently regulated by the observer.

Results.—The first substance selected for observation was calcium sulphate with a trace of manganese as activator which had been previously prepared for the study of the thermo-luminescence.[2] After exposure to x-rays for thirty minutes or more about 100 mg. of the white powder was placed in tube *P* and a like amount in *P'*. A slow stream of gas from the ozonizer was passed through tube *P* for about fifteen minutes after which a portion of the material from *P* and *P'* was placed in two circular depressions about 1 cm. in diameter countersunk to a depth of 0.3 mm. in a thin copper plate side by side (Fig. 2), forming equal layers of the ozone treated and of the non-treated material. Upon a hot metal plate over a small circular electric stove at a temperature well below the red heat, the characteristic rich green thermo-luminescence of this substance

was promptly developed but the brightness of the control in all of the numerous trials was *greatly in excess of that of the sample which had been exposed to ozone*. Under the conditions prevailing in our apparatus an exposure of 15 minutes was found to be adequate to produce maximum quenching.

Equally striking results were obtained by the above method with:

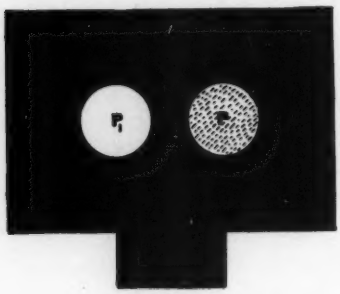


FIGURE 2

Treated sample (P) and untreated sample P_1 on heating plate.

Cadmium phosphate (a highly phosphorescent and thermo-luminescent material made by the late W. S. Andrews).

Calcites from Langban, England, and from Franklin Furnace, N. J.

Other materials in which the effect was less striking but unmistakable were:

Fluorites known to contain traces of rare earths.

Zinc sulphide (a strongly phosphorescent preparation from Rousseau which had been rendered neutral by heating before being subjected to the action of x-rays).

Calcium sulphide (a Lenard and Klatt "phosphor" known as No. 3, which contained calcium sulphide with bismuth as an activator).

Sodium chloride (common salt which had been turned yellow by x-rays and which glowed with a fine lemon-yellow thermo-luminescence).

In short, every substance capable of thermo-luminescence by treatment with x-rays, upon which we experimented, suffered diminution of brightness after exposure to ozone.

Thermo-luminescence by x-rays has been shown[3] to consist, usually, of a preliminary flash-up, bright but of comparatively short duration, followed by a more prolonged glow differing from the first in color. The quenching effect of ozone appeared to be more nearly complete with respect to this first glow. Thus in the case of calcium sulphate the thermo-luminescence of which passes over from a rich yellow-green to blue-green, the diminished glow of ozonized samples approached the latter hue from the very beginning of their thermo-luminescence.

The explanation is that in the spectrum of thermo-luminescence there are usually at least two broad overlapping bands, the relationship of which to time and temperature is such that upon the application of heat there are progressive changes in the color of the glowing substance. It is as though the momentary effect came from the reduced form of the substance produced by x-rays, and gave way to the more persistent glow as

oxidation progressed. In the presence of ozone the first stages of oxidation would be completed before heat was applied and the thermo-luminescence would be that of later stages.

The reducing action of x-rays has been abundantly demonstrated by the researches of Paul Bayley[4] upon the coloration of certain salts. Thermo-luminescence occurs when the reduced product is comparatively stable at room temperature but oxidizes spontaneously and rapidly when heated. Diminution of the brightness of thermo-luminescence after exposure to ozone, as in the experiments above described, is to be ascribed to partial re-oxidation during exposure to that agent at ordinary temperatures. If such oxidation occurs with sufficient rapidity one might look for an emission of light and in fact we were able to detect a feeble glow as the ozone passed over the material in tube *P* of our apparatus, discernible only in complete darkness and with eyes rendered sensitive by previous exclusion of light. Some of our materials, notably the calcium sulphate above mentioned, were sufficiently luminescent at room temperature, even when a considerable time had elapsed subsequent to their treatment with x-rays, to emit a glow perceptible in complete darkness. In all such cases a slight enhancement of the glow during the passage of the oxygen containing O_3 could be perceived.

Cathode Rays.—Since cathode rays are also known to render various substances thermo-luminescent[5] we repeated our experiments on calcium sulphate, substituting cathodo-bombardment in vacuo for exposure to x-rays. The effect of ozone was in every observable respect the same and doubtless the same explanation applies, for, as has been established by D. T. Wilber,[6] cathode rays expel oxygen from this class of substances.

Luminescent solids clearly fall into two classes: those which are, so to speak, self contained and under excitation suffer little or no loss of oxygen but only a temporary local displacement or separation with an immediate spontaneous return. To this class belong many crystals such as kunzite, the ruby, etc. In the vacuum tube under cathodo-bombardment they continue to glow indefinitely even after all oxygen has been removed from the surrounding atmosphere.

In the other class are various oxides such as CaO, ZrO_2, SiO_2 and Al_2O_3. In the cathode tube they glow brightly during the early stages of evacuation, but diminish in luminescence as oxygen is more completely removed from the tube, are more or less discolored due to partial reduction of their surfaces and soon become almost or entirely unresponsive to bombardment. When oxygen is restored to the tube their power of luminescence returns. This effect was first described in 1921[7] and the explanation developed in the recent experiments by Wilber, just cited.

The action of ozone upon thermo-luminescence described in the foregoing paragraphs is a special case in the consideration of the nature of

luminescence and of the relation to it in general of ozone. The matter which has been treated at some length in a recent paper[8] may be very briefly stated as follows:

(1) Excitation, whether by corpuscular bombardment or the action of radiation of the shorter wave-lengths, sets free oxygen and the resulting glow is due to the return of the oxygen to a state of combination. Excitation, then, is a reducing process. In favor of this view we have the following facts:

(*a*) The glow may be prevented by the removal of the oxygen as rapidly as it is formed, as in certain cases of cathodo-bombardment[9] or by the presence of a reducing agent such as hydroquinone, as in solutions of the fluorescent organic dyestuffs.[10]

(*b*) Excitation may be omitted altogether where reduction has previously taken place by ordinary chemical processes, if free oxygen be provided from without as in the glow of phosphorus or of the other light-giving oxidations known to the chemist, or in the luminescence of vital organisms, or in the experiments of Ramsey with fluorescein, etc., just cited, or in cando-luminescence.

(2) An activator is known to be necessary to the production of luminescence in so many cases that it is reasonable to assume that it is an essential factor. The activator may indeed be regarded as the source of the free oxygen and the solvent itself may have no part in the dual reduction-oxidation process to which the glow is due.

(3) The oxygen set free by excitation at low temperatures is in part at least in active form, since ozone is always associated with the reducing action of ultra-violet radiation, etc. In cando-luminescence where the glow is produced by the hydrogen flame active oxygen is likewise present. Since ozone is the agent by means of which the luminescent glow may be produced in the *absence of excitation*, and since it or an equivalent form of active oxygen occurs where the glow is *due to excitation*, it is reasonable to infer that it is a necessary agent in luminescence, upon which the presence or absence of the glow depends.

[1] Wick, F. G., *Phys. Rev.*, (2) **24**, p. 272; also Wick and Gleason, *J. O. S. A.*, **9**, p. 639, 1924.

[2] Wick, loc. cit.

[3] Wick, loc. cit.

[4] Bayley, *Phys. Rev.*, (2) **24**, p. 495, 1924.

[5] Wick, loc. cit.

[6] Wilber, D. T., *J. O. S. A.*, **20**, 1930 (abstract of paper read before Opt. Soc., 1929).

[7] Nichols and Wilber, *Phys. Rev.*, (2) **17**, p. 707, 1921.

[8] Nichols and Boardman, *J. O. S. A.*, **20**, 1930.

[9] Wilber, D. T., loc. cit.

[10] See the work of Ramsay.

GEOMETRICAL OPTICS IN ABSORBING MEDIA

BY PAUL S. EPSTEIN

CALIFORNIA INSTITUTE OF TECHNOLOGY

Read before the Academy, November 19, 1929

1. *Introductory.*—The problem of propagation of rays in a medium of variable absorption and refraction has acquired some practical importance in radiotelegraphy. The radio waves penetrate into the conducting and absorbing layer of the upper atmosphere[1] and describe there a certain path. The question was brought to my attention by Dr. B. Van der Pol of Eindhoven (Holland) who asked me to give a formula by which the shape of a ray in the conducting layer could be computed. The solution given in this paper is not restricted to electromagnetic rays, but applies, equally well, to waves and rays of other types. With respect to the radio waves, one of our results is that the rays, after going up, do not bend down again and do not ever come back to earth. The rays, observed by the receiving station, as coming from the conducting layers are not primary but secondary waves produced in it by reflection. The writer is preparing a second paper dealing with the problem of space reflection in an inhomogeneous absorbing medium.

Before tackling the problem of absorption, let us reduce to its simplest terms the ordinary theory of geometrical optics in *transparent* media. As starting point, we may choose the *Fermat* principle, according to which light (sound, etc.) travels from a point A to a point B in the shortest time:

$$S = c^{\cdot}\int_A^B dt = \text{minimum.} \tag{1}$$

If the medium has a variable index of refraction n, the velocity of propagation is given by $v = c/n$, where c represents the velocity in a standard medium (vacuum in the case of light). For this reason, we may multiply the element dt of our integral by

$$v^2n^2/c^2 = (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)n^2/c^2 = 1, \tag{2}$$

obtaining

$$S = \int L dt, \tag{3}$$

with

$$L = n^2v^2/c = (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)n^2/c \tag{4}$$

2. *Rays in Transparent Media.*—Our problem is now formally equivalent with a Hamiltonian dynamical problem having the Lagrangean function (4). The Lagrangean equations, representing the equations of a ray, evidently become

$$2\frac{d}{dt}(n^2\dot{x}) - \frac{\partial n^2}{\partial x}\dot{x}^2 = 0, \text{ etc.} \tag{5}$$

For the solution of these equations, we may proceed on the lines of the Hamiltonian theory. We introduce a "momentum" by the equations $p_x = \partial L/\partial \dot{x}$, etc., or

$$\bar{p} = \bar{v}n^2/c, \tag{6}$$

and a Hamiltonian

$$H = -L + \bar{v}.\bar{p} = p^2n^2 = v^2n^2/c^2. \tag{7}$$

From this point of view, equation (2) can be written as

$$H = (p_x^2 + p_y^2 + p_z^2)n^2 = 1; \tag{8}$$

it is the analog to the energy integral of dynamics with a special choice of the energy constant. We make use of a further relation from Hamilton's theory

$$\bar{p} = \nabla S, \tag{9}$$

and, substituting this into equation (8), we obtain Hamilton's partial differential equation in the form

$$(\nabla S)^2 = n^2, \tag{10}$$

which in geometrical optics is called "*equation of the iconal.*" An integral of this equation will have the form

$$S = S(x, y, z, \alpha_1, \alpha_2) + C, \tag{11}$$

where α_1 and α_2 are two non-additive integration constants. The physical meaning of the quantity S in geometrical optics is the "*phase of the wave.*" The equation

$$S(x, y, z, \alpha_1, \alpha_2) = \text{const.} \tag{12}$$

represents, therefore, a surface of equal phase. Written in differential form it becomes

$$\frac{\partial S}{\partial x}dx + \frac{\partial S}{\partial y}dy + \frac{\partial S}{\partial z}dz = 0, \tag{12'}$$

showing that its characteristics

$$\frac{dx}{\partial S/\partial x} = \frac{dy}{\partial S/\partial y} = \frac{\partial z}{\partial S/\partial z}, \tag{13'}$$

which are equivalent to equations (5), are at the same time the equations of the normal trajectories: *the rays are everywhere normal to the wave surfaces.*

For obtaining the integrated form of equation (5) or (13′) we may use the *Jacobi* method which consists in differentiating (11) with respect to α_1,α_2 and equating the partials to two new constants β_1, β_2:

$$\frac{\partial S}{\partial \alpha_1} = \beta_1 \quad , \quad \frac{\partial S}{\partial \alpha_2} = \beta_2. \tag{13}$$

This is the final expression for the equations of a ray in a transparent medium.

3. *Absorbing Media.*—In the accurate theory of wave optics, the state of a medium is controlled by the equation

$$\nabla^2 u + k^2 n^2 u = 0 \tag{14}$$

($k = 2\pi/\lambda$, where λ is the wave-length in the standard medium). In an absorbing medium, the index of refraction is complex $n^2 = \kappa + i\sigma$. If n is constant (homogeneous medium) the simplest solution of equation (14), corresponding to a plane wave, is

$$u = \exp.\ [i(a_1x + a_2y + a_3z) - (b_1x + b_2y + b_3z)], \tag{15}$$

where the coefficients a and b must satisfy the single condition

$$(a_1 + ib_1)^2 + (a_2 + ib_2)^2 + (a_3 + ib_3)^2 = n^2 = \kappa + i\sigma. \tag{16}$$

The meaning of expression (15) is that we obtain the state of the optical field in a point x, y, z, at the time t, if we multiply (15) by the time factor $\exp(-ikct)$ and take the real part of the product. We see that we have to distinguish here two families of surfaces: the planes $a_1x + a_2y + a_3z =$ const. are *surfaces of equal phase*, the planes $b_1x + b_2y + b_3z =$ const. are *surfaces of equal amplitude*, or of equal intensity. The two families are in general distinct and do not coincide.

In the more general case of an inhomogeneous medium, n is a function of the coördinates. The method of geometrical optics applies to the case when n changes appreciably only in distances which are long compared with the wave-length λ. In a small volume, whose linear size is but a few λ, n remains practically constant and it is permissible to use the expression (15) considering in it a, b to be functions of x, y, z. We arrive so at the fundamental expression of a wave in geometrical optics of absorbing media

$$\begin{cases} u = \exp(ikS_1 - kS_2) = \exp(ikS) \\ S = S_1 + iS_2 \end{cases} \tag{17}$$

S_1 and S_2 are two functions of x, y, z. Equation $S_1(x, y, z) =$ const. represents the surfaces of equal phase, $S_2(x, y, z) =$ const. represents the surfaces of equal amplitude.

We derive a partial differential equation for the complex quantity S by substituting in the usual way[2] the approximate solution (16) into equation (14) and neglecting terms of lower degree in k than k^2. We obtain

$$(\nabla S)^2 = n^2. \tag{10'}$$

Formally this equation is identical with (10), but both S and n are now complex numbers. This relation, therefore, contains two simultaneous equations for the determination of S_1 and S_2. An integral of the form (11) determines both the surfaces of equal phase

$$S_1 = Re\ S(x, y, z, \alpha_1, \alpha_2) = \text{const.}, \tag{11'}$$

and those of equal intensity

$$S_2 = Im\ S(x, y, z, \alpha_1, \alpha_2) = \text{const.}, \tag{11''}$$

where the symbols *Re* and *Im* denote the real and imaginary parts of the expressions following them

We have seen that the variation principles (1) and (3) are mathematically equivalent with (10). However, it is hardly desirable to use them in the theory of absorbing media, because the quantities t and v lose their simple physical meaning. In fact, the phase S is now complex and, therefore, can no longer be interpreted as time. It is, therefore, best to base the whole theory on equation (10') of the iconal and not to make use of Fermat's principle.

4. *Rays in an Absorbing Medium.*—In a transparent medium, the only set of surfaces characterizing a wave motion are the surfaces of constant phase. The direction of a ray is always normal to these surfaces. We have seen in the preceding section that waves in absorbing media are characterized by two sets. Since the time of Poynting's theorem, we understand under a ray the curve in which the energy travels. In the case of electromagnetic waves, we can determine the instantaneous direction of the energy flow by means of Poynting's vector. To fix our ideas, let us consider the two-dimensional case, in which the index of refraction n and the optical field are independent of the coördinate z. Let us, further, assume that the electric vector is parallel to $z (E_x = E_y = 0)$. For E_z, we substitute the approximate expression (16) of geometrical optics which we write in its real form

$$E_z = \exp(-kS_2) \cos k(S_1 - ct). \tag{18}$$

According to Maxwell's equations the magnetic field is given by $H_z = 0$,

$$\mu H_x = \exp(-kS_2)\left[\frac{\partial S_1}{\partial y} \cos k(S_1 - ct) + \frac{\partial S_2}{\partial y} \sin k(S_1 - ct)\right],$$

$$\mu H_y = -\exp(-kS_2)\left[\frac{\partial S_1}{\partial x} \cos k(S_1 - ct) + \frac{\partial S_2}{\partial x} \sin k(S_1 - ct)\right].$$

The components of Poynting's vector are

$$P_x = -cE_zH_y, \quad P_y = cE_zH_x, \quad P_z = 0 \tag{19}$$

and the direction of it is given by the ratio

$$P_y/P_x = -H_x/H_y \tag{20}$$

In the case of transparent media, $S_2 = 0$ and this ratio becomes independent of time. Poynting's vector has in every point of space a constant orientation, and this is the reason why we can compute the curves of energy flow. It is not so *in the case of absorbing media*, because here Poynting's vector oscillates: we cannot deduce the energy path from this vector and must look for some other method for doing this.

From the consideration of a single infinite wave, it is impossible to determine either the velocity or the direction of the flow of energy, as we have no way of identifying the individual energy elements. To determine the velocity, Lord Rayleigh used a group of wave frequencies. We shall try to use a group of waves of different orientations, in order to determine the path of the energy. This can be made in the most convincing way by considering the properties of a limited pencil of rays.

5. *A Pencil of Rays.*—In order to produce a pencil of rays, the experimental physicist uses a diaphragm. The theoretical one can do the same by means of Huyghens' principle. The expression for the field of a pencil in air, which, originating in an infinitely removed source, has passed through a rectangular diaphragm is[3]

$$u = u_0 \int_{-a}^{+a} \int_{-b}^{+b} \exp\,[ik(z' + \alpha x' + \beta y')]\, d\alpha d\beta. \tag{21}$$

In this expression α and β are small numbers whose squares and higher powers are neglected. This is permissible if we observe the pencil at a large distance from the diaphragm (Fraunhofer diffraction). We see that u represents the superposition of an infinitely large number of plane waves whose wave normals include different small angles with the z' axis.

The choice of coördinates in expression (21) is a special one: z' has the direction of the axis of the pencil, but by a proper adjustment of the complex constant u we can choose the origin at any distance from the diaphragm. Let us generalize our equation by a rotation of coördinates going over to the new system x, y, z. If we denote the substitution angles $\cos(xx') = l_{xx'}$, etc., the phases S_0 of the individual waves in air will have the expression

$$\begin{aligned} S_0 = {} & (\alpha l_{xx'} + \beta l_{xy'} + l_{xy'})x + (\alpha l_{yx'} + \beta l_{yy'} + l_{yz'})y \\ & + (\alpha l_{zx'} + \beta l_{zy'} + l_{zz'})z. \end{aligned} \tag{22}$$

Now we let our pencil pass into a second absorbing homogeneous medium, selecting as the surface of discontinuity the xy-plane. Each of the con-

stituent plane waves of integral (21) will be refracted into a wave of the type (15) in the second medium. We can, therefore, make use of the well-known rules for the refraction of plane waves,[4] according to which the coefficients of x and y remain unchanged, while the coefficient of z must be computed from relation (16). This leads to the following formula for the second medium

$$\begin{aligned} S' &= (\alpha l_{xx'} + \beta l_{xy'} + l_{xz'})x + (\alpha l_{yx'} + \beta l_{yy'} + l_{yz'})y \\ &+ \sqrt{n^2 - (\alpha l_{xx'} + \beta l_{xy'} + l_{xz'})^2 - (\alpha l_{yx'} + \beta l_{yy'} + l_{yz'})^2}\, z. \end{aligned} \tag{23}$$

The expression for the pencil in the second absorbing medium will be

$$u = u_0 \int_{-a}^{+a} \int_{-b}^{+b} \exp(ikS')\, d\alpha d\beta. \tag{24}$$

Strictly speaking, not only the phase S_0 but also the amplitude u_0 is changed in the process of refraction. But since our plane waves impinge all at about the same angle, α and β being very small, this will involve only effects which are immaterial for our purpose.

6. *Equation of a Ray.*—To be consistent, we have to expand expression (23) with respect to α and β and to retain terms of zero and first order only. If we substitute the result into (24), we can carry out the integration and we obtain

$$u = u_0 \exp(ikS) \frac{\sin kA}{kA} \frac{\sin kB}{kB}. \tag{25}$$

$$\begin{cases} S = l_{xz'}x + l_{yz'}y + (a_3 + ib_3)z, \\ A = l_{xx'}\left(x - \dfrac{l_{xz'}}{a_3 + ib_3} z\right) + l_{yx'}\left(y - \dfrac{l_{yz'}}{a_3 + ib_3} z\right), \\ B = l_{xy'}\left(x - \dfrac{l_{xz'}}{a_3 + ib_3} z\right) + l_{yy'}\left(y - \dfrac{l_{yz'}}{a_3 + ib_3} z\right). \end{cases} \tag{26}$$

The axis of the pencil is there where the absolute value of the complex expression (25) has its maximum. We find the maximum in the routine way, multiplying (25) by its complex conjugate, differentiating the product with respect to x and y, and making the two partials equal to nothing. The result of these operations is

$$x/l_{xz'} = y/l_{yz'} = Re(z/(a_3 + ib_3)).$$

As

$$l_{xz'} = \partial S/\partial x \quad , \quad l_{yz'} = \partial S/\partial y \quad , \quad a_3 + ib_3 = \partial S/\partial z,$$

the last equation can be also put into the form

$$Re\left[\frac{x}{\partial S/\partial x} = \frac{y}{\partial S/\partial y} = \frac{z}{\partial S/\partial z}\right]. \tag{27}$$

This is the equation of a ray in a homogeneous medium. If the medium is inhomogeneous, we make use of the principle of geometrical optics, according to which we still can apply the results found for homogeneous media for an infinitesimal length of the ray. Let us use curvilinear orthogonal coördinates given by the line element

$$ds^2 = Udu^2 + Vdv^2 + Wdw^2. \tag{28}$$

Our equation becomes

$$Re\left[\frac{Udu}{\partial S/\partial u} = \frac{Vdv}{\partial/S\partial v} = \frac{Wdu}{\partial S/\partial w}\right]. \tag{29}$$

It must be noted that, in view of our derivation, this equation applies only to a choice of coördinates in which

$$\partial S_2/\partial u = \partial S_2/\partial v = 0. \tag{30}$$

This can be always reached by selecting the surfaces of equal intensity (S_2 = const.) as one set (w = const.) of our orthogonal mesh system. In most cases occurring in applications conditions (30) are satisfied automatically. One usually considers systems where absorption and refraction are produced by the same physical cause, so that the surfaces κ = const. and σ = const. coincide. Thus they give the appropriate coördinates for the solution of equation (10′) and, at the same time, they turn out to be identical with S_2 = const.

As equations (29) represent the characteristics of (10′) in curvilinear coördinates, we again may use the Jacobi method and obtain

$$Re\ \partial S(u, v, w, \alpha_1, \alpha_2)/\partial\alpha_1 = \beta_1\ ,\ Re\ \partial S(u, v, w, \alpha_1, \alpha_2)/\partial\alpha_2 = \beta_2, \tag{31}$$

provided that constants α_1, α_2 are real.

7. *Examples of Application.*—Let us consider a stratified medium, i.e. a medium in which the index of refraction is a function of the coördinate z only; let, further, our rays lie in the (x, z)-plane. The problem then becomes two-dimensional and we have from (10′) the equation

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 = n^2(z), \tag{32}$$

whose solution is

$$S = \alpha x + \int\sqrt{n^2 - \alpha^2}\,dz. \tag{33}$$

To fix our ideas, we shall apply this to the case of radiowaves. Since these waves start in the non-absorbing lower layers of the atmosphere α is real. For the rays, we obtain (31)

$$Re \int \frac{dz}{\sqrt{n^2 - \alpha^2}} = \frac{x}{\alpha}, \tag{34}$$

or in differential form

$$\frac{dx}{dz} = Re \frac{\alpha}{\sqrt{n^2 - \alpha^2}}. \tag{35}$$

In the case of the index of refraction decreasing with height, there is a fundamental difference between transparent and absorbing media. (1) *If the medium is transparent,* the ratio dz/dx decreases with height until it becomes equal to zero for $n = \alpha$. The rays describe an arc, reaching their maximum height at this elevation, and bend back to earth. This is the well-known phenomenon of total reflection. (2) *If the medium is absorbing* ($n^2 = \kappa + i\sigma$), the ratio dz/dx has always the same sign. The ray goes up monotonically without ever bending round and coming down. The rays which come down from absorbing layers are not totally, but partially reflected waves. As long as the absorption is small ($\sigma \ll \kappa$), there is not much difference between this case and the preceding one: The primary rays become almost horizontal and produce a strong reflection only in this part of their path, while the secondary, reflected rays have here almost the same direction as the primary ones.

It seems, however, that in the upper layers of the atmosphere, the conduction and absorption are quite considerable.[5] A further investigation is, therefore, desirable dealing with the intensity and path of reflected rays.

Formula (35) contains the law of refraction for rays passing from a homogeneous transparent medium (n_0) into a homogeneous absorbing medium ($n^2 = \kappa + i\sigma$). If the angle of incidence is β_0 (angle between ray and normal to the surface of discontinuity in transparent medium), the geometrical meaning of α is $\alpha = n_0 \sin \beta_0$. The corresponding angle in the second medium is given by

$$tg\beta = \alpha \sqrt{\frac{\kappa - \alpha^2 + \sqrt{(\kappa - \alpha^2)^2 + \sigma^2}}{2(\kappa - \alpha^2)^2 + 2\sigma^2}}. \tag{36}$$

This formula shows that the deviation from the ordinary law of refraction is noticeable only then when the absorption in a thickness equal to one wave-length is appreciable.

[1] The conducting layer of the upper atmosphere is usually referred to as the "*Heavi-*

side layer." The priority in postulating it belongs, however, to Prof. A. E. Kennelly whose paper on this subject appeared in March, 1902 (*Electrical World and Engineer*, p. 473). *Heaviside's* article was published in December, 1902 (*Encyclopaedia Brittannica*, Vol. 33, p. 215) and was written in June, 1902, according to his own testimony (*Electromagnetic Theory*, Vol. 3, p. 335).

[2] E. g., A. Sommerfeld und I. Runge, *Ann. Physik*, 35, p. 290, 1911.

[3] E. g., M. v. Laue, *Enzyklopädie Math. Wissenschaften*, Vol. 5, p. 424, 1915.

[4] W. Wien, *ibid.*, p. 130.

[5] P. O. Pedersen, *Proc. Inst. Radio Engineers*, 17, p. 1750, 1929.

RECENT PROGRESS IN THE DUAL THEORY OF METALLIC CONDUCTION

BY EDWIN H. HALL

JEFFERSON PHYSICAL LABORATORY, HARVARD UNIVERSITY

Communicated December 7, 1929

In 1921 I published[1] a "Summary" of the numerical results tentatively arrived at by means of the dual theory of conduction, especially the values of the "characteristic constants" of the various metals dealt with. Recently I have revised these numerical estimates, availing myself of some new data furnished by the International Critical Tables and applying the results of more mature reflection upon the problem with which I am engaged. The present paper may be regarded as a revision of the "Summary" above referred to. It is not a complete statement of all the features and applications of the dual theory in its present state of development, but it undertakes to show what measure of success this theory has attained in dealing numerically with the relations between electrical conduction, thermal conduction, the Thomson effect and the Peltier effect, in eighteen metals, including two alloys.

The "characteristic constants" of a metal, according to this theory, are the z, q, C, C_1, C_2, λ_c' and s contained in the following equations:

$$n = zT^q, \tag{1}$$

$$(\kappa_f + \kappa) = C + C_1 t + C_2 t^2 \tag{2}$$

and

$$\lambda' = \lambda_c' + skT. \tag{3}$$

In (1) n is the number of free electrons contained in unit "free space" within a metal at temperature T. By "free space" is meant space in which the free electrons can move as gas particles—that is, the whole volume of the metal less that part from which the free electrons are excluded by the atoms and ions. How much the "free space" differs from the total volume I cannot at present undertake to say, but the distinction

here made, analogous to that which is familiar to all who have dealt with the van der Waals equation for an imperfect gas, seems to be a valid and necessary one.

In equation (2) κ_f means that part of the total specific electric conductivity, κ, which is due to the free electrons, the other part, which will be called κ_a, being due to the "associated" electrons which are supposed to progress through the metal by passing from an atom to an adjacent ion without becoming free, and so without sharing the energy of thermal agitation. The t indicates temperature above 0°C.

In equation (3) λ' means the amount of energy, in ergs, required to dissociate an atom within the metal, thus producing a free electron, with a full complement of thermal energy, and a positive ion, both products remaining within the solid metal. The k is the Boltzmann gas-constant.

It is quite possible that, in order to meet all the conditions of the case, some of these so-called "characteristic constants," the q and the s, for example, must be supposed to vary slightly with temperature, but provisionally, and for the range of temperature between 0°C. and 100°C., this being the whole range with which I am at present concerned, they are to be treated as constants.

These three equations, with the interpretations here put upon them, embody the fundamental assumptions on which the dual theory of conduction is based. Our immediate concern is to derive numerical values for each of the "characteristic constants," except z, for every one of the metals with which we here have to do.

To determine the values of six unknown quantities we need to have six independent equations connecting these quantities. Three of the needed equations I have obtained from a study of the Thomson effect. These three equations, which are too long to be reproduced here, were first derived in a paper published by myself in 1920 (*Proc. Nat. Acad. Sci.*, **6**, pp. 139–154, equations (14), (15), (17)) and were revised the same year (*Ibid.*, **6**, pp. 613 and 614).

The fourth equation to be used for our present purpose is the expression for thermal conductivity, θ, in the terms of the dual theory. Such an expression is given as equation (20) in a paper[2] published in 1926, where its derivation also is given. But I now use a slightly different form, which is

$$\theta = \left(\frac{k}{e}\right)^2 \kappa_f \left[\frac{\kappa_a}{\kappa}(s + q - 1.5)\left(\frac{\lambda'_t}{k} + sT\right) + 2T\right]. \qquad (4)$$

Here the only new letter is e, which indicates the electron charge, 1.59×10^{-20}, a quantity which is to be taken with the positive sign in all of my discussions of the dual theory. The new feature of this revised expression for θ is the term $2T$, introduced in order to take account of that part of thermal conduction which is due to the mere process of diffusion of the

free electrons along a temperature gradient, as distinguished from their general tendency to drift down this gradient. The contribution thus made to θ, which contribution I shall call θ_l, is

$$\theta_l = \left(\frac{k}{e}\right)^2 . \kappa_f . 2T. \tag{5}$$

The subscript l refers to Lorentz, who, taking the whole of electrical conduction to be due to the free electrons, wrote[3] the equivalent of

$$\theta = \left(\frac{k}{e}\right)^2 . \kappa . 2T$$

It seems probable that, in order to get a complete expression for θ, it would be necessary to add to the second member of equation (4) a term representing the contribution made by the agitation of the atoms in their lattice arrangement; but for the present I have undertaken to get on as well as I can without this, not knowing how great it should be. It is possible that the somewhat unsatisfactory character of the results, presently to be shown, which my calculations give for θ at 100°C. is due to this omission.

The fifth equation to be used in evaluating the constants is derived from one which I published in 1921,[4] but the full significance of which I have only recently appreciated. This latter equation is equivalent to

$$\frac{n_\alpha}{n_\beta} = \exp\,[(\lambda'_\beta - \lambda'_\alpha) \div kT]. \tag{6}$$

Substituting for n and λ' in (6) by use of (1) and (3), then writing the equation in logarithmic form, differentiating with respect to T and finally multiplying by T, I get

$$q_\alpha - \frac{(\lambda'_c)_\alpha}{kT} = q_\beta - \frac{(\lambda'_c)_\beta}{kT}. \tag{7}$$

Now in obtaining (7) q, λ'_c and s have all been treated as constants. But evidently, if q and λ'_c are strictly constant through any range of temperature, $(\lambda'_c)_\alpha$ must equal $(\lambda'_c)_\beta$ and q_α must equal q_β. It is possible, however, by making $(\lambda'_c \div k)$ small and making q_α nearly, though not exactly, equal to q_β, to make equation (7) very nearly true and very useful for our present purpose.

What is the number which all the values of q should approximate? This question is not difficult to answer. In my previous (1921) estimation of the characteristic constants I had not discovered the restriction imposed by equation (7) and my values of q ranged from 1.2, for iron, to 1.6 for certain other metals. The mean, however, was not far from 1.5, a number which occurs repeatedly in the evaluating equations derived

from the Thomson effect and always in the combination $(1.5-q)$. In fact, as a rule, an invariable rule for cases in which the Thomson effect is simply proportional to T, a metal having, according to my sign convention, a positive Thomson effect was found to have a value of q greater than 1.5, while a negative Thomson effect implied a value of q less than 1.5. This number, then, was plainly indicated as the one about which the values of q should cluster. Acting accordingly I have now found, as the smallest value for q, 1.49, in iron, and as the largest 1.505, in cobalt and in constantan. Meanwhile the value of $(\lambda_c' \div kT)$ at 0°C. has fallen from a former maximum of 2.91, for bismuth, to a maximum of about 0.02, for magnesium. These changes have not only enabled me to comply with, or at least give respectful attention to, the condition suggested by equation (7); they have had a most happy result in my dealing with the numerical data of the Peltier effect.

The sixth and last of the equations needed for evaluating the constants is an expression, which I published in 1921,[5] for the Peltier effect between two metals α and β. It is

$$\Pi_{\alpha\beta} = \left(\frac{\kappa_f}{\kappa}\right)_\beta \lambda_\beta - \left(\frac{\kappa_f}{\kappa}\right)_\alpha \lambda_\alpha. \qquad (8)$$

$\Pi_{\alpha\beta}$ means, with my convention as to algebraic signs, the amount of heat, in ergs, absorbed by the unit quantity of electricity, $(10^{20} \div 1.59)$ electrons, in going from metal α to metal β. I shall in all cases take bismuth as the metal β, because by so doing I have only positive values for $\Pi_{\alpha\beta}$. For my Peltier-effect numerical data, as for my Thomson-effect numerical data, I depend entirely upon the work of Bridgman as published[6] in his too-little celebrated paper on *Thermo-Electromotive Force, Peltier Heat and Thomson Heat under Pressure.*

After the characteristic constants for bismuth have been found, so that the term $((\kappa_f \div \kappa)\lambda)_\beta$ in equation (8) is known, I undertake to find for every other metal such a set of constants that this equation will give the observed value of $\Pi_{\alpha\beta}$ at 0°C. But naturally the question arises how the constants for bismuth itself can be determined. Might we not vary the α constants and vary those of β also in such a way that equation (8) would still give a value agreeing with the observed value of $\Pi_{\alpha\beta}$ at 0°C.? Yes, but unless a particular set of constants is used for β, or bismuth, equation (8) will not give the correct, the observed, values of $\Pi_{\alpha\beta}$ at both 0° and 100°C. This criterion, which I have arrived at almost accidentally after considerable experimentation, appears to be entirely satisfactory. After I had found a set of constants for bismuth and a set of constants for any other, α, of my metals, which sets of constants complied with the other conditions imposed and, when used in equation (8), gave the observed values of $\Pi_{\alpha\beta}$ at 0° and at 100°, I had no further trouble

with the Peltier effect. Using the same set of bismuth constants and taking any other metal α, I found that, when I had obtained for this metal a set of constants complying with the other conditions and satisfying equation (8) at 0°, the same set of constants satisfied equation (8) at 100°.

TABLE 1

VALUES OF THE "CHARACTERISTIC CONSTANTS"

METAL	C	$C_1\times10^6$	$C_2\times10^8$	q	s	$\lambda_c'+k$	δ_0	δ_{100}	$\tau \div T$
Al	0.1083	−0.47	7.33	1.499	4.75	2.8	0.11	0.15	$-0.016 + 0.006t$
Bi	0.338	−91.5	0	1.499	4.06	4.0	0.10	0.13	−3.2
Co	0.1545	197	0	1.505	4.60	2.6	0.11	0.15	7.8
Cu	0.0737	−17.4	0	1.499	6.45	4.5	0.15	0.21	−0.966
Gold	0.0686	−15.6	0.8	1.499	6.95	3.6	0.16	0.22	$-0.934 + 0.001t$
Iron	0.0380	24.1	35.2	1.490	8.49	3.4	0.20	0.27	$1.78 + 0.0516t$
Mg	0.0800	−0.15	0	1.499$_3$	6.36	5.5	0.15	0.21	−0.008
Mo	0.0606	−68.8	11.9	1.495	7.31	3.3	0.17	0.24	$-4.33 + 0.015t$
Ni	0.1787	103.3	0	1.502	4.00	2.9	0.09	0.13	3.56
Pd	0.0934	66.7	0	1.503	6.12	3.4	0.14	0.20	3.52
Pt	0.0856	48.7	0	1.5025	6.36	3.7	0.15	0.21	2.67
Silv.	0.0727	−15.2	0	1.4995	6.60	2.5	0.16	0.21	−0.864
Tl	0.0742	4.68	2.99	1.4995	6.60	3.2	0.16	0.21	$0.268 + 0.0034t$
Tin	0.0742	2.27	0	1.5001	6.85	3.2	0.16	0.22	0.134
Tung.	0.0578	−46.6	0	1.498	8.50	3.0	0.20	0.27	−3.41
Zn	0.0716	17.33	0	1.501	6.63	3.9	0.16	0.21	0.99
Con.	0.1423	143.7	0	1.505	6.41	3.6	0.15	0.21	7.94
Man.	0.0527	−10.2	−4.14	1.501	9.40	4.2	0.22	0.30	$-0.828 - 0.0067t$

In table 1 are assembled the values found for the "characteristic constants" of the eighteen metals, including the two alloys constantan and manganin, with which I have dealt.[7] The numbers for $(\lambda_c' \div k)$ are here carried to one place only of decimals, though in some of my calculations I found it desirable to use two decimal places for this quantity. The δ_0 and δ_{100} columns give approximately, in volts, the ionizing potentials within the several metals at 0°C. and at 100°C., respectively, these values being found from those of λ' in equation (3). The last column of the table gives, with my sign convention, the values of the Thomson effect, as found from the work of Bridgman above referred to. For each metal the value of τ gives the amount of heat, in ergs, absorbed by the unit quantity of electricity, $(10^{20} \div 1.59)$ electrons, in going from a place of temperature $(T - 1/2)$ to a place of temperature $(T + 1/2)$ within the metal. It is to be noted that, in all cases where τ is a one-term quantity, q is greater than 1.50 when τ is positive and less than 1.50 when τ is negative.[8]

This last column of table 1 and columns 1, 2, 3, 5, 7 and 9 of table 2 contain the data which the dual theory of conduction endeavors to coordinate, through the "characteristic constants" of table 1 and the equa-

tions in which these constants are used. Unfortunately, the data which these columns give with respect to any one metal have in no instance all been derived from study of the same specimen of that metal. Thus I take all my thermoelectric data, as I have already said, from the work of Bridgman, but he did not measure specific conductivity, either electrical

TABLE 2

VALUES AT 0° AND AT 100°C. OF κ, θ AND Π, IN ABSOLUTE MEASURE

METAL	$\kappa_0 \times 10^6$	$\kappa_{100} \times 10^6$	$\theta_0 \times 10^{-5}$ OBS.	CALC.	$\theta_{100} \times 10^{-5}$ OBS.	CALC.	$(\Pi_{\alpha\beta}) \times 10^{-4}$ OBS.	CALC.	$(\Pi_{\alpha\beta})_{100} \times 10^{-4}$ OBS.	CALC.
Al	415	289	203	202	209	193	202.0	202.1	263.0	262.8
Bi	9.4	6.5	8.4	8.3	7.7	7.7	0	0	0	0
Co	110.6	81.0	69.2	69.0	69.2	76.3	155.9	155.9	172	171.3
Cu	643	450	388	390	381	365	210.8	211.2	279.6	280.0
Gold	450	322	296	294	296	282	211.1	211.1	279.8	279.3
Iron	112.4	69.3	61.9	61.7	59.8	59.7	247.3	247.4	309.7	309.5
Mg	242	174	155	154.3	147	151.4	203.0	203.3	265.4	265.5
Mo	228	159	146	146.2	139.4	126.9	219.3	218.9	300.9	300.2
Ni	106	74.6	58.6	58.2	57.7	59.6	155.1	154.8	186.7	186.1
Pd	98.5	74.7	67.4	67.2	72.0	74.0	188.1	188.6	232.0	232.3
Pt	102.6	74.0	69.5	69.4	73.2	72.0	194.7	195.1	244.1	244.2
Silv.	668	475	419	417	412	398	210.2	210.4	278.3	278.2
Tl	60.9	41.3	39	38.8	40.5	36.3	207.7	207.9	270.2	270.1
Tin	95.5	66	65.7	65.4	60.5	61.8	203.8	203.6	266.0	265.4
Tung.	194	147	160	159.2	158	152.2	207.5	207.8	284.3	284.4
Zn	181	127	113	112.5	111.3	110.3	211.5	211.7	273.3	273.0
Con.	20.4	20.4	22.0	21.9	29	32.5	108.3	108.4	106.4	105.9
Man.	22.7	22.7	20.8	20.8	26.4	27.7	206.9	206.6	275.1	274.3

or thermal, and so for much of my data I have had to depend on information given in the International Critical Tables, referring to the work of other men using different material, and in many cases the information thus obtained is very unsatisfactory for my purpose. Nevertheless, I have, in the procedure described in this paper for estimating the values of the constants, been obliged to treat the body of data which I have settled upon for any one metal as authoritative and self-consistent. Thus the values of τ given in table 1 and those of κ_0 and κ_{100} given in table 2 have gone into my calculations as absolutely true, and accordingly the values of the constants which have resulted from these calculations must be consistent with these data. The values of θ and of Π given in columns 3, 5, 7, and 9 of table 2 have been used in a somewhat different manner. Instead of incorporating these values directly into my formulas, I have sought, by a tentative process, for a set of constants that would, when used in equations (4) and (8), respectively, yield values of θ and Π agreeing satisfactorily with the so-called "observed" values of these quantities.

As regards θ_0, $(\Pi_{\alpha\beta})_0$ and $(\Pi_{\alpha\beta})_{100}$, this undertaking has been entirely successful. In no case does the "calculated" value of one of these quan-

tities differ from the "observed" value by so much as one per cent, and even closer agreement could, I think, have been obtained by taking more pains in the selection of the characteristic constants. Anyone who is interested can with little trouble verify my "calculated" results for one or more metals by putting the needed constants, given in table 1, and the observed values of κ_0 and κ_{100}, given in table 2, into equations (4) and (8), with due regard to equations (2) and (3).

The "calculated" values of θ_{100}, on the other hand, do not agree very well with the "observed" values, as the latter are given in table 2. The mean of the calculated values is here about 2.8 per cent smaller than the mean of the "observed" values, and in individual cases the difference is much greater than this, about 10 per cent, for example, in the case of cobalt and molybdenum.

Is this partial failure sufficient reason for giving up the whole theory of conduction, in spite of its apparent success in other respects? I think not, though of course I am not an impartial judge in this matter. There are three possibilities which seem to me worth considering in this situation. First, it may be that my neglect, already acknowledged as unfortunate though for the present necessary, of the thermal conductivity due to agitation of the atoms is the cause of this defect regarding θ_{100}. Second, perhaps a thoroughly successful dual theory of conduction will have to take account of a possible variation, with temperature, in the value of q or of s or of both. Third, perhaps more accurate data, obtained for each metal from the same specimen or from specimens of the same quality, would eliminate the difficulty now encountered. For example, in the case of cobalt, already mentioned as one of the metals for which the discrepancy regarding θ_{100} is exceptionally large, I have put down 69.2×10^5 as observed value of both θ_0 and θ_{100}. But in fact all the information I have found in the I. C. T. concerning θ in cobalt is that someone found the value to be 69.2×10^5 at 30°C. in a specimen with 0.24% C, 1.4% Fe, 1.1% Ni, 0.14% Si. From this I have, by use of the Wiedemann-Franz rule, $\theta \alpha \kappa T$, found 69.2×10^5 as the "observed" value of both 0° and 100°; for it so happens that in cobalt, according to the observations of Bridgman on the specimen he used, the value of κ is almost exactly inversely proportional to T. As to molybdenum, in which also the discrepancy between observed and calculated θ_{100} is large, the I. C. T., volume 5, p. 218, give the value of the Wiedeman-Franz "constant" as 3.08 at 0° and 3.17 at 100°, the "theoretical" value for all metals being 2.23.

In fact, when I consider the many doubts attaching to the body of data with which I have dealt in this investigation, I am quite content to have an appreciable amount of discrepancy between the "observed" and the "calculated" values exhibited in columns 5 and 6 of table 2. Complete agreement between calculated and observed quantities everywhere would

arouse, even in myself, a suspicion as to the soundness of my argument and the value of my results, for I should fear that I had somehow been reasoning in a circle and begging the questions at issue.

It is to be noted that my "observed" θ_{100} values were not used at all in the determination of the "characteristic constants." Accordingly, if I had for every metal a body of data which I regarded as correct and self-consistent, my theory, if sound and complete, should enable me to find by calculation the value of θ_{100} for each metal independently of any use on my part of the observed value at the temperature in question. Thus the findings of the theory could be subjected to a genuine test. It seems possible, though I do not care to prophesy to that effect, that with full and satisfactory data concerning the electrical and thermal properties of metals the dual theory might account for those departures from the Wiedemann-Franz law which are so marked and at present so mysterious. I cannot help regretting that some fraction of the research energy and talent and material resources now devoted to attacks upon the constitution of the atoms cannot be turned to a study of the familiar but little understood properties of metals in their ordinary condition.

There is something more to be said concerning equation (6), the validity of which does not depend on the soundness of the dual theory of electric and thermal conduction as set forth by myself. Apparently this equation must be accepted by anyone who admits the existence of free electrons, with full thermal energy, within metals and admits also the conclusion, drawn from the observed equality of "stopping potentials" for photo-electrically emitted electrons, that the total energy of an associated, or valence, electron is the same in all metals in contact with each other at the same temperature. For with these conditions equation (6) is merely the Boltzmann law of equilibrium distribution of free electrons between the two joined metals, α and β, $(\lambda'_\beta - \lambda'_\alpha)$ being the amount of energy required to take an electron from the free state in α to the free state in β by the following reversible process: association of a free electron in α with an ion of α to form a neutral atom, passage of the electron from the atom of α into union with an ion of β at the junction of the metals, escape of the electron from this union into the free state in β.

What the dual theory, if correct, contributes here is the means of determining λ'_α and λ'_β numerically and so finding the ratio $(n_\alpha \div n_\beta)$ in definite numerical terms for any two of the metals dealt with. It is true that equation (6) is primarily derived for metals in actual contact with each other, but is evident that no one of the quantities n_α, n_β, λ'_α, λ'_β, occurring in this equation, can be dependent to any appreciable extent on the circumstance of contact or lack of contact between the metals. So equation (6), if it holds for two metals in contact, holds for the same two metals when separate, provided the two have the same temperature.

From the values of s and $(\lambda_c' \div k)$ given in table 1 we can, by use of equation (3), get the value of λ' for any one of our metals at any temperature from 0° to 100°C. Table 3 gives values of $(n_\alpha \div n_\beta)$, at $T = 273°$ and at $T = 373°$, for every metal dealt with, β being in every case tungsten. Tungsten is given prominence here because of the leading part it plays in studies of electronic emission and in other experiments bearing upon the value of n. According to the numbers given in this table the ratio $(n_\alpha \div n_\beta)$ is larger than 1 in every case except that of the alloy manganin. In the case of bismuth it is very large, 84.7, and in that of nickel still larger, 89.1. For copper the ratio given is about 7.7, and for constantan, composed of about 60% Cu and 40% Ni, it is about 8.1, a value lying between the ratios for nickel and copper, respectively, but much nearer the latter. For manganin, composed of about 84% Cu, 12% Mn and 4% Ni, the value of n appears to be especially small.

TABLE 3

VALUES OF $(n_\alpha \div n_\beta)$, β BEING IN ALL CASES TUNGSTEN

METAL α	AT 0°C.	AT 100°C.	METAL α	AT 0°C.	AT 100°C.
Aluminum	42.5	42.5	Nickel	89.1	89.1
Bismuth	84.7	84.7	Palladium	10.81	10.81
Cobalt	49.4	49.4	Platinum	8.49	8.49
Copper	7.69	7.76	Silver	6.68	6.68
Gold	4.71	4.71	Thallium	6.68	6.68
Iron	1.01	1.01	Tin	5.21	5.21
Magnesium	8.41	8.49	Tungsten	1	1
Molybdenum	8.93	8.93	Zinc	6.49	6.49
Constantan	8.09	8.09	Manganin	0.40	0.41

Of course, no great amount of confidence can be placed in values arrived at as these have been, but they seem to me worth recording, and it is possible that they may have some influence in suggesting the course of future experiments for direct evidence[9] as to the magnitude of n.

The experiments of Miss Colpitts[10] indicated, somewhat uncertainly, 2×10^{17} per cu. cm. of the metal as the lower limit of n in gold and 9×10^{15} per cu. cm. of the metal as the lower limit in tungsten. The ratio of these two n's is about 22. The ratio given in table 3 for gold and tungsten, per cu. cm. of "free space" in each metal, is 4.7.

NOTE.—At the Solvay Conference of 1924, which dealt with the subject of electric conduction in metals, the question arose as to whether the Boltzmann distribution law in its simplest form holds between n, the number of "free" electrons, and n_0, the number of un-ionized atoms, per cu. cm. of the metal. Q being the amount of energy required to ionize an atom within the metal, the law in question can be stated, tentatively, thus:

$$n = n_0 \epsilon^{-\frac{Q}{kT}}. \qquad (A)$$

It was in fact so stated by O. W. Richardson in the following inquiry

(p. 128 of the *Conference Report*):[11] "Does the formula [given above] hold in the theory of quanta?"

A number of the conferees present confidently expressed the opinion that it did hold. Others, including Lorentz and Bridgman, expressed a doubt. A reader of the whole *Report* would be likely to get the impression that the Conference approved the use of the equation, in the form given above, as governing the amount of ionization within a metal. One member, speaking in support of this use, said, "The Boltzmann equation applies even in cases where one can apply the rules of quanta. It is thus that Einstein made use of it in 1917 in his derivation of the law of Planck."

The paper of Einstein here referred to is printed in the *Physikalische Zeitschrift*, vol. **18**, pp. 121–128. The following passage occurs on p. 122: "According to the quantum theory a molecule of a given kind can, apart from its orientation and motion of translation, take only a discrete series of conditions $Z_1, Z_2, \ldots Z_n, \ldots$, the internal energy of which shall be $e_1, e_2 \ldots e_n \ldots$ If molecules of this sort belong to a gas of temperature T, the relative frequency W_n of the condition Z_n is given by the formula

$$W_n = p_n \epsilon^{-\frac{e_n}{kT}}, \qquad (5)$$

corresponding to the canonical distribution of states in statistical mechanics. In this formula k $(= R \div N)$ is the familiar Boltzmann constant, p_n a number, independent of T, characteristic of the molecule and of its nth quantum condition, which can be defined as the 'statistical weight' of this condition. The formula (5) can be derived from the Boltzmann principle or by a purely thermodynamic method. Equation (5) is the expression for the widest generalization of the Maxwellian distribution law of velocities."

It is to be observed, however, that Einstein's eq. (5), when applied to the two quantum conditions m and n, gives

$$\frac{W_m}{W_n} = \frac{p_m}{p_n} \exp\left[-\frac{e_m - e_n}{kT}\right]. \qquad (B)$$

There is nothing in Einstein's paper to show that $(p_m \div p_n)$ must be 1. Accordingly it would seem that Richardson's question should be answered in the negative, unless the formula which he gave is to be amended so as to read

$$n = \frac{p_n}{p_{n_0}} n_0 \epsilon^{-\frac{Q}{kT}}, \qquad (C)$$

where $(p_n \div p_{n_0})$ is a factor to be determined.

[1] *Proc. Nat. Acad. Sci.*, **7**, pp. 98–107(1921).

[2] *Physic. Rev.*, **28**, pp. 392–417(1926).

[3] See p. 67 of his *Theory of Electrons*.

[4] *Proc. Nat. Acad. Sci.*, **7**, p. 63, eq. (8).

[5] *Ibid.*, **7**, p. 63, eq. 10.

[6] *Proc. Amer. Acad. A. S.*, **53** (1918), pp. 269–386.

[7] I have not included cadmium in this list of metals. Bridgman studied cadmium as he did the eighteen other metals, but the value which he found for the Thomson effect in it is about ten times as great as that indicated by the observations of Fleming and Dewar (*Phil. Mag.*, **40**, 5th series, 1895). Bridgman expresses the opinion that the specimen of cadmium which he studied may have been in an unstable condition, between two allotropic forms.

[8] Although I have arrived quite independently at the number 1.50 as the mean value of q in metals, I have already (*Nat. Acad. Sci. Proc.*, **15**, p. 507) called attention to the fact that Richardson in deriving his T^2 emission formula from the "classical kinetic theory" assumed the number of free electrons per cm.^{3} of a metal to be proportional to $T^{1.5}$.

[9] I have discussed evidence bearing upon this magnitude in a recent paper, "On Electrons That Are 'Pulled Out' from Metals," *Proc. Nat. Acad. Sci.*, **15**, 241–251 (1929).

[10] *J. Frank. Inst.*, **206**, Oct., 1928.

[11] Gauthier-Villars, Paris, 1927.

TWO CONTRADICTIONS IN CURRENT PHYSICAL THEORY AND THEIR RESOLUTION

BY F. S. C. NORTHROP

DEPARTMENT OF PHILOSOPHY, YALE UNIVERSITY

Communicated November 29, 1929

When Einstein examined the electro-magnetic theory of Lorentz at the beginning of this century, he found the principle of relativity for Galilean frames of reference, and the principle of the absolute velocity of light in vacuo, to be experimentally valid, yet contradictory. The necessity of rejecting the doctrine of absolute space and time for the principle of the relativity of simultaneity, in order to escape this contradiction, led him to the discovery of the special theory of relativity.

It is the purpose of this paper to show, by precisely the same type of argument, that this theory and the general theory which came out of it have given rise to two new contradictions in current physical theory which can be escaped only by introducing a radical but essentially simple amendment to the traditional atomic theory.

The Difficulty Concerning Atomicity and Motion.—The first contradiction will be demonstrated by establishing three propositions. (1) Atomicity is an inescapable fact. (2) Atomicity and the motion which it involves necessitate the existence of a referent in addition to the microscopic particles. (3) No such referent exists, according to current scientific theory.

1. *Atomicity is an inescapable fact.* The first argument for the atomic theory was given in Greek science by the pre-Socratic philosophers. They

demonstrated, by reasoning similar to that which drove Einstein to the theory of relativity, that one cannot admit the existence of the two obvious extensive facts of stuff and change, and not accept the atomic theory, without involving one's self in a contradiction. Since these two facts cannot be denied, the same logic which makes us accept the relativity theory requires that we accept the atomic theory.

The technical investigations of modern science have supported this conclusion with many independent lines of evidence, so great in quantity as to be practically overwhelming. Only a few items in a long list need be referred to. This list begins with Newton's acceptance of the atomic theory and with LaPlace's statement of the laws of his mechanics in terms of it, and ends with the discovery of the atomic character of electricity and energy. In between is Count Rumford's discovery and Maxwell's development of the kinetic theory of heat, the modern science of chemistry, the discontinuity which statistical theories suggest, the evidence indicating the physico-chemical character of living things, Sir J. J. Thomson's isolation and Millikan's determination of the charge of the electron, and astronomical theory, beginning with LaPlace and continuing through Jeans, which conceives of the primordial state of stellar evolution as atomic in character.

It would appear that nothing more needs to be said. Recently, however, two types of theory have been proposed which suggest the reduction of atomicity to a non-atomic basis. I refer to Eddington's doctrine that mathematical forms, rather than physical atoms, are fundamental,[1] and to current developments in wave mechanics.

Three things need to be said concerning Eddington's thesis. Firstly, it rests for its validity upon the theory of relativity which in turn presupposes an atomic theory. The relativity theory has its origin and foundation in the electro-magnetic theory of Lorentz, and this is an atomic or electron theory. Secondly, Eddingtons' extremely mathematical theory derives its validity from a generalization beyond the theories of Einstein and Weyl which is not universally accepted. In fact, Einstein has unequivocally repudiated it.[2] Thirdly, even if it be accepted, it requires one to admit the primacy of physical instead of mathematical categories. For an examination of the postulates in terms of which Eddington states this extremely mathematical theory[3] reveals that they involve statements concerning "displacements" "carried by parallel displacement" from point to point. Certainly the carrying presupposes motion, and that which is carried cannot be a mathematical relation. Quite obviously, as Einstein explicitly asserts in the definition of his geometry, it is a measuring rod. Now, measuring rods and motion are physical things, and involve the existence of a plurality of objects, which is intelligible only on an atomic basis. It becomes evident, therefore, that atomic categories are really primary.

It was this necessity of defining geometry in terms of physical measuring rods and their behavior which led the pure mathematician Riemann, who invented the mathematics of the general theory, to predict, long before Einstein's discovery, that space would be found to be conditioned by matter.

The continuous and apparently non-physical emphasis in wave mechanics also turns out upon examination to be an argument for, rather than against, the atomic theory. Two considerations will suffice to make this clear. Firstly, all theories, no matter how mathematical may be their emphasis, use Plank's constant as a fundamental idea. This constant involves an atomic or discontinuous theory of energy. Furthermore, it arose from a new study of the foundations of statistical theory which rests upon the conception of finite rather than infinitesimal units in terms of which possible permutations are to be defined. In other words, Planks' constant is essentially a discontinuous or atomic idea. Secondly, even the attempt in current quantum theory, to conceive of the atom as a singularity in a continuum, breaks down when one makes this rather vague notion the least bit explicit. For in order to give it any fertility, it is necessary to determine the potential and kinetic energy of the supposed singularity, and this involves the conception of it as an actual particle to which the inverse square law applies.

Thus, even in those cases in which the atomic theory appears to be discarded, it is really a fundamental and basic factor. We are forced, therefore, by the facts to regard our first proposition as established. Atomicity is an inescapable fact.

2. *Atomicity and motion necessitate the existence of a referent in addition to the microscopic particles.* The proof of this proposition was given in Greek science by Parmenides. It was accepted by Newton and suggested by him at the beginning of the Principia. The rejection of absolute space by the theory of relativity has led many to suppose that the proof in question is invalid. A reëxamination of it indicates that this is not the case. Current scientific theories will be examined and found to support this conclusion.

Consider the proof as Parmenides gave it. Change, he said, must be due to generation or to motion. After demonstrating that it cannot be due to generation if this is a physical universe, he proved that it cannot be due to motion if the physical stuff is conceived as nothing more than one substance or many microscopic particles. For motion requires that stuff be transmitted from where it is to where it is not, and there can be no "where-it-is-not" if nothing but the stuff to be transmitted exists.

Nor can change be due to nothing but many particles and their motion—for two reasons. Firstly, the motion of many particles calls for a "where-it-is-not," as much as the motion of one. A difficulty is not removed

by multiplying it many times. Secondly, there cannot be many particles if nothing exists but the stuff of which they are supposed to be made. For before there can be many parts of stuff there must be something to divide one part of stuff from another and this is impossible if nothing but the stuff to be divided exists.

The essential point in this argument is not so much the need for an intervening medium, as the need of a basis for distinguishing between one atom and another. In an atomic theory which regards the atoms as constituted of the same kind of material, the category of stuff gives only the respect in which the atoms are identical or one; it cannot prescribe the respect in which they are numerically different or many. Stated positively, this means that the individuality of a given atom cannot be defined in terms of its properties, but must be expressed in terms of its unique relation to some referent common to all the atoms. If nothing but the stuff of the atoms exists there can be no such referent, and hence atomism is impossible.

It may be said, therefore, that one can no more admit the validity of the atomic theory without accepting the existence of a referent other than the microscopic particles, than one can accept the postulates of Euclidean geometry without accepting its theorems. As Newton said, "It is from their essence or nature that they (things) are places."[4] To think one without the other is impossible.

It is to be noted that the last argument takes care of those who would attempt to define atomic motion in terms of the relation of the atoms to each other. For even the notion of relative motion involves the existence of many atoms, and the manyness, wholly apart from the motion, is impossible unless a referent other than the atoms exists.

Let us not evade the full consequence of this conclusion. It means that any physical theory of relative motion involves a more basic atomic theory of absolute motion. For, let me repeat again, relative motion involves many substances and the manyness is meaningless apart from their relation to a common referent other than them. Verified theories of contemporary science support this conclusion.

Consider the kinetic theory of gases, and a particular case—the molecules of a gas in a container. It is to be noted that when we talk about the motion of these molecules we do not mean their relation to the container. For we believe, and it is part of our theory that we must believe, that the molecules will continue to move if the container is destroyed. This could not be the case if their motion had no meaning apart from the existence of the container. Moreover, it is part of statistical theory to maintain that the container, as well as the phenomena within, is to be defined in molecular terms. Eventually, therefore, we are driven back behind all containers to the totality of molecules of the whole of nature. Since we define all molar objects in terms of their ultimate molecular motion, we cannot use

them to define what we mean by molecular motion. An examination of what we mean by motion as exhibited in the phenomenon termed the Brownian movement leads to the same conclusion.[5]

These considerations remind us that there are two types of valid scientific theory in contemporary science. The one, exemplified by Newtonian and Einsteinian mechanics, states all scientific observations in terms of relations to chosen reference frames; the other, exemplified in all statistical and kinetic atomic theories, makes use of atomic entities in absolute motion relative to a common referent and escapes the relativity attaching to a use of reference frames, without the recourse to transformation equations. In the latter type of theory, motion never refers to a molar body or to a frame chosen by a scientist; it is conceived as occurring independently of any such choice.

This is clearly revealed in the astronomical theory of Jeans. This theory is particularly suitable for our purposes because it combines relativity and atomic conceptions within a single scientific theory. His appeal to the relativity theory to account for the tremendous radiation of energy from the sun is well known. This necessitates his rejection of the doctrine of absolute space. Nevertheless, when he lays down the elemental principles of his theory, he refers to original nature as a collection of atoms moving in space.[6]

Now, it is to be noted that this space is not relative space; nor will the facts which force him to this theory permit him to regard it as relative space. For relative space has no meaning apart from reference frames which involve molar bodies, and at the beginning of stellar evolution molar bodies do not exist. Nor can he mean that the atoms are moving relative to some member of their group to which a scientist chooses to refer them, for at the stage in question conditions are not very propitious for the existence of scientists. Furthermore, if he meant that the atoms are moving relatively to each other, he would be forced to refer to them, not as moving *in* space, but as giving rise, because of their motion, to a system of relationships between each other which *is* space, and this would require that he introduce some other referent to indicate what he means by motion. It must never be forgotten that a relative theory of space must define space as a relation between objects; it can never define it as a system of relations or a container relative to which or in which atoms move. The plain fact of the matter is that the kinetic atomic theory with which Jeans begins requires a referent for motion which can be neither a molar body nor a chosen particle, and since the discarded absolute space is the only one which science knows, he uses it.

The result, however, is a contradiction. For his acceptance of the theory of relativity necessitates the rejection of the absolute space upon which he founds his theory. But he is not to be singled out in connection with this

fallacy. Every scientist commits it who talks about atomicity and motion while accepting the relativity theory in its current interpretation.

The point to note now, however, is that necessary current scientific theories support Parmenides' proof of our second proposition, which is that atomicity, and the motion which it involves, necessitates the existence of a referent other than the microscopic particles.

3. *No such referent exists*, according to current scientific theory. The traditional referent, absolute space, disappeared with the acceptance of the special theory of relativity. The general theory eliminates space-time as a possible referent. For it indicates that the metric of space-time is conditioned by matter and its motion. Hence, we would involve ourselves in a vicious circle if we used space-time to define what we mean by matter and motion.

With this establishment of our third proposition the contradiction in current scientific theory is made evident. Both the atomic theory and the relativity theory are supported by factual evidence, yet the latter theory as currently interpreted, denies the existence of the referent which the former theory necessitates. In other words, a situation similar to the one which drove Einstein to the discovery of the special theory is at hand. Two verified theories exist which contradict each other.

Only one course is possible in such a circumstance. Current scientific theory must be modified to fit the facts. A conception must be discovered which will permit us to accept both of these theories without contradiction. The procedure used by Einstein indicates how this new conception is to be discovered.

He argued that a contradiction between verified principles proves that the presupposition in our theory which gives rise to the contradiction must be false. He proceeded, therefore, to determine what the presupposition is, and to replace it by its negate.

In his case, this meant the rejection of the doctrine of absolute time for the doctrine of the relativity of simultaneity. In our case it means that we must cease to identify the non-existence of absolute space with the non-existence of a referent for atomic motion. Stated positively, this means that a new referent must replace the discarded absolute space. Only in this way can we accept the atomic theory and the doctrine of the relativity of space which the relativity theory necessitates, without involving ourselves in a contradiction. Thus, just as the contradiction which Einstein found in electro-magnetic theory drove him to the discovery of the special theory of relativity, so the contradiction which we have found in current physical theory drives us to the discovery that a new referent for atomicity and motion must exist.

The Difficulty Concerning Metrical Uniformity.—The question arises concerning what this new referent is. This brings us to a second contradiction

in current physical theory which arose out of a difficulty first indicated by Whitehead.[7] We shall attempt to demonstrate its existence by establishing three propositions. (1) The metrical properties of space are conditioned by the motion and distribution of matter. (2) The space of this universe possesses an approximately constant metrical uniformity which extends over macroscopic distances. (3) Matter as conceived by current physical theory is incapable of producing this uniformity.

1. *The metrical properties of space are conditioned by the motion and distribution of matter.* This proposition is a consequence of the general theory of relativity. As Einstein has said: "According to the general theory of relativity the metrical character (curvature) of the four-dimensional space-time continuum is defined at every point by the matter at that point and the state of that matter. Therefore, on account of the lack of uniformity in the distribution of matter, the metrical structure of this continuum must necessarily be extremely complicated."[8] Recent astronomical observations have supported previous ones[9] to reconfirm this conclusion.

2. *The space of this universe possesses an approximately constant metrical uniformity which extends over macroscopic distances.* This must be admitted because it is presupposed in the measurement of astronomical distances.

A few considerations will make this clear. Consider a distance here on the earth. The measurement of its length presents no difficulty. One merely takes a standard rod and notes how many times it must be laid down end to end to cover the distance in question. However, the location of an astronomical distance is such as to make a direct determination of its length impossible. The difficulty is met by establishing a relation of equality between the astronomical distance and a local terrestrial one to which the rod can be directly applied. To establish this relationship, it is necessary to appeal to geometrical principles. Moreover, if the results are to be valid these principles must apply to the intervening space which relates the two distances in question.

If the metrical structure of this space varied, two difficulties would occur. Firstly, the geometrical principles to which we appeal at one time would not be those which we used at another time. Were this true it would follow that the values arrived at by an observation in one century should not combine with those determined in another to make sense. In short, the values in question would be incommensurable. Secondly, we should not know what geometrical principles to use in a single given observation until we had determined the distribution of matter in the intervening region in question at the time. But we cannot determine the distribution of matter in astronomical space without measuring astronomical distances. Thus we find ourselves in the peculiar predicament of not being able to make a single astronomical measurement of distance until we have made a large

number of such measurements. The experience of astronomers does not confirm this consequence. It must be admitted, therefore, that at least an approximately constant metrical uniformity extends over macroscopic distances in space.

3. *Matter as conceived in current scientific theory is incapable of producing this metrical uniformity.* One of the most notable consequences of the general theory of relativity is its doctrine that the metric of space changes with the distribution of matter. Whitehead and others pointed out that it is difficult to understand how measuring is possible if the metrical variability which this entails exists.

To meet this and other difficulties, and to explain why we have not discovered this metrical variability before, Einstein was forced to state his general theory in terms of a conception of nature as a whole, and to assert that the metrical variability is a local microscopic factor in a general macroscopic uniformity.[8] There seems to be little doubt but that this is the case. Certainly the measurement of astronomical distances would not be possible if the general uniformity were not present. Also, recent and previous observations in astronomy indicate that the metrical variability exists.

Many have supposed that this leaves everything in a satisfactory form. Quite the contrary is the case. The difficulty on which Whitehead put his finger has been merely shifted into another form in which it becomes all the more acute. Instead of asking why microscopic variability is so rare, we now have to ask why macroscopic uniformity is so obvious. In other words, the rule rather than the exception to it presents a difficulty. For if the metric of space is conditioned by matter, as the evidence for the general theory forces us to admit, it follows from our current theory of matter that the general macroscopic metrical uniformity should not exist. In short, exactly the opposite of what the facts indicate and Einstein suggests should occur.

This becomes clear the moment one notes that a relational theory of space which is conditioned by matter means that the metric of space is defined in terms of relations between physical objects; ultimately, this means, in terms of relations between atoms. In such a theory, metrical variability exists when the relations between the atoms change; and metrical uniformity when they do not.

Now, it is a verified doctrine of physics, and an essential principle of any physical theory which would account for change, that these atoms are in motion. This means that the relation between them is changing and hence that metrical variability should be the general rule. The second law of thermo-dynamics and the modern doctrine that matter is indifferent to order or design is an expression of this same fact. It is precisely because of this that our predecessors were unable to refer the spacial characteristics of nature to matter.

Einstein attempted to avoid this consequence by defining the required metrical uniformity in terms of our traditional theory of matter.[8] As one would expect, this necessitated that he regard matter as practically at rest. Certainly it is only upon this assumption that the relations between the parts of matter will not change, and metrical uniformity can result. But this assumption is not in accord with our knowledge of the kinetic character of matter. The evidence for the kinetic atomic theory makes this way out of the difficulty impossible.

Heisenberg's principle of uncertainty[10] supports this conclusion. For the contingency which it introduces into atomic behavior necessitates not merely that metrical heterogeneity and variability should be the rule, but also that no law governing the variation can be found. We must conclude, therefore, that matter, as currently conceived, is incapable of producing the metrical uniformity which exists.

The second contradiction in current physical theory is now demonstrated, for the three propositions which constitute the demonstration are established. Let us bring them together. A constant metrical uniformity exists which extends over macroscopic distances. The metric of space is conditioned by matter. Matter as conceived by current scientific theory is incapable of producing such a metric. In other words, current scientific theory asserts that space is conditioned by matter, in its doctrine of mensuration and space; and denies this thesis in its doctrine of matter.

The Resolution of the Two Contradictions.—From such a contradiction only one escape is possible. The last of our three propositions must be rejected, since it is the only one of the three which rests upon theory rather than fact. This means that we must admit that our traditional theory of the capacity of matter to produce order is false. As Einstein indicated, in connection with the evidence which drove him to the relativity theory, when facts contradict, the presupposition in our traditional theory which produces the contradiction must be rejected and replaced by its negate.

Consider what this means in our case. We know that a metrical uniformity extending over macroscopic distances exists, and the facts behind the general theory of relativity necessitate that we hold that the metric of space is conditioned by matter. We know also that matter, as traditionally conceived, cannot produce this uniformity. It follows therefore that our traditional theory of the capacity of matter to produce order is false and must be amended to meet new evidence.

We have noted that matter is atomic in character. Hence this second contradiction concerning space, as well as the first contradiction concerning atomicity and motion, drives us to the conclusion that an amendment must be made in our traditional atomic theory.

Moreover, a consideration of the two major characteristics of the metric of this universe is sufficient to indicate precisely what the required amend-

ment must be. This metric combines two different characteristics. In macroscopic regions it exhibits metrical uniformity and constancy, and in microscopic regions it exhibits metrical variability.

The material basis for the latter characteristic is already at hand in our traditional atomic theory. For, if we mean by the metric of space the relation which joins the atoms to each other, and by metrical variability a change in this relationship, it follows from the kinetic and contingent character of the behavior of the microscopic atoms that metrical variability must exist. It follows also that the material basis for a constant uniform relatedness or metric cannot be found in the microscopic atomic entities. The most that can be expected from them is an occasional appearance of order and constancy which will be but an incidental factor within the general relational variability.

We discover, therefore, that some other basic entity in addition to the microscopic particles must exist in this universe. Otherwise the macroscopic metrical uniformity would not exist. In fact, we have but to note what is required to enforce a general constant macroscopic metrical uniformity upon the variable metrical relatedness of the microscopic particles to discover precisely what the properties of this new entity are.

Firstly, it must be physical. For if it is to cause the microscopic atomic entities, and the molar objects into which they compound, to take on a constant uniform order which their motion alone cannot produce, it must change the direction of their motion. This calls for the presence of an external force, which only a physical object can provide.

Secondly, this physical entity must congest and completely surround all the microscopic atomic entities of the whole of nature. Otherwise they would be merely crowded out into some other referent for their motion, and macroscopic metrical uniformity would not exist.

Thirdly, this additional physical entity must be an atom, rather than a compound substance. Otherwise some referent other than it would be needed to provide a basis and meaning for the distinction between any one of its parts and another, and the old difficulty over atomicity would recur.

We have but to bring together these three required characteristics to discover the existence of a large macroscopic atom, spherical in shape and hollow in its interior except for its inner field, which contains and congests all the microscopic atoms of the traditional atomic theory.[11] Moreover, we seem to have no alternative but to say that the facts not merely suggest but necessitate the existence of this atom. For, as we have indicated, one cannot deny its existence without denying certain facts or involving one's self in two contradictions.

Note how this radical but essentially simple addition to our traditional atomic theory meets both of the difficulties to which we have previously

referred. In providing a referent for atomicity and motion it enables us to accept the atomic theory and the rejection of absolute space which the relativity theory entails. For it permits us to define space-time as a relation between objects without losing a meaning for the atomicity and motion of those objects. And by introducing a fixed spherical physical form, surrounding the microscopic particles, which congests them sufficiently to preserve a general constant relational uniformity, while not congesting them so much as to prevent their motion and the resultant local variable relatedness, it provides us with a metric, completely conditioned by matter, which must exhibit local microscopic metrical variability within a general macroscopic metrical uniformity. In short, it gives us a kinetic theory of matter which reconciles the macroscopic metrical uniformity which measuring requires, with the microscopic metrical variability and the complete dependence of space-time upon matter which the general theory of relativity necessitates. By surrounding the electrons and protons of the whole of nature with one large finite spherical macroscopic atom, atomic physics and relativity physics are brought within a single consistent physical theory. This theory, we shall henceforth refer to as the macroscopic atomic theory.

It appears that traditional science was right in insisting that atomism necessitates the existence of a referent in addition to the microscopic particles, but wrong in identifying this referent with space. It appears also that traditional relativity physics has been right in maintaining that space and time are relative, but wrong in denying the existence of any referent for atomic motion and in maintaining that all motion is relative. It appears also that the static character of astronomical matter, to which Einstein referred in his cosmological reflections, exists. But it finds its basis in the congesting pressure from without of the macroscopic atom, rather than in the conception of all matter as at rest. Thus, the relatively static character of stellar matter and the macroscopic metrical uniformity of space is reconciled with the overwhelming evidence for the kinetic character of the microscopic atoms.

Consequences and Unique Verifications.—The macroscopic atomic theory has several important consequences. Those bearing on the nature of space-time, and the three-dimensional physical interpretation of the theory of relativity which it provides, were given in the first paper on this theory.[11] Others will be indicated here.

Firstly, the universe is finite. This conclusion is not new in relativity physics. However, our argument to it is original, for it presupposes nothing more than the kinetic atomic theory, measuring, and the general theory of relativity. No assumptions concerning boundary conditions or the stability of the universe are necessary.

Secondly, there is a physical ether. But, instead of being an absolute thing independent of atomic matter, it is the varying complex field which

results from the compounding of the inner field of the macroscopic atom with the many fields and central charges of the microscopic electrons and protons.[11] This compound field is the electro-magnetic-gravitational field of which we and our earth form a part, and in which we are imbedded. It is to this compound field and its potential distribution that the tensor equations of relativity theory refer. This changing compound field may also be called space-time. When so considered it is to be remembered, however, that the temporal dimension is a mere abstraction from motion.[11] This identification of the ether with this changing compound field provides a physical medium for the transmission of electro-magnetic waves.

Thirdly, the discovery of an ether-drift would not be incompatible with the theory of relativity. For the part of the ether which we call the earth might or might not move with a different velocity than the larger part of the ether which immediately surrounds the earth. If it did so move, a drift or a whole series of drifts might be detected. This would prove nothing, however, concerning the absolute velocity of the earth, since it is not known whether the part of the changing ether which immediately surrounds the earth is at rest or in motion relatively to the macroscopic atom. It is because molar objects are imbedded and move in this complex changing ether that their motion is relative; the principle of isolation which is the basis of Newton's first law of motion no longer holds. It is because the microscopic atoms produce the ether and space-time and move in the macroscopic atom that their motion is absolute.

Fourthly, neither the geometry of the general theory nor of the generalizations of Eddington[3] and Weyl[12] is adequate. They fail to provide for metrical uniformity over more than infinitely short displacements. The macroscopic atomic theory calls for metrical uniformity over macroscopic finite distances. In addition, it requires an inclusion of electromagnetic potentials. Otherwise the compound field could not be conditioned by the electrons and protons of electro-magnetic theory. Also, it must be Riemannian. Otherwise the variable potential distribution which the motion of the microscopic atoms necessitates would not be possible, for a variable g_{ik} distribution can only be stated in a law of an invariant form, by a geometry of the Riemannian type. It is a geometry with these three characteristics that Einstein has developed in his unitary field theory. His latest achievement may be regarded, therefore, as a confirmation of our theory. For a detailed development of this point, see another paper[13] *of these Proceedings*.

Fifthly, astronomical phenomena which are inexplicable on traditional principles seem to be a necessary consequence of the macroscopic atomic theory. Traditional astronomical theory has been based upon the microscopic atoms and their motion. Since our theory includes the traditional microscopic atoms it possesses all the deductive fertility of traditional

theory. In addition, the influence of the macroscopic atom is present. This adds to traditional forces a rotational force resulting from the deflection of the microscopic particles from a rectilinear course by the physical spherical form of the macroscopic atom. To be sure, this deflection would occur originally, only with those microscopic atoms which are moving near the spherical shell of the macroscopic atom, but eventually it is reasonable to suppose that it would be transmitted through the whole interior of nature. We are thus provided with a force of pure rotation in addition to the forces of traditional principles. Astronomical evidence indicates that this force is required. I refer to what Jeans terms "the characteristic equiangular spiral shape of arms" which is present in "at least nine tenths of the spiral nebulae."[14] He has shown that this calls for a force of pure rotation. The attempt of Professor E. W. Brown of my university to account for this in terms of traditional forces necessitates the hypothesis of a very complicated distribution of matter.[14] The situation seems to be analogous to the problem of the motion of the orbit of Mercury in Newtonian mechanics and Leverrier's hypothesis of the undiscoverable plant Vulcan. Jeans concludes that a force of pure rotation is called for, which traditional principles cannot provide.[14] If we may be permitted to reason from purely physical and qualitative considerations, it may be said that this constitutes a unique verification of the macroscopic atomic theory. It may be noted that rotational motion has always presented difficulties for traditional physics.

Sixthly, our theory indicates that all motion is not relative. The motion of the microscopic atoms in the macroscopic atom is absolute. This means that the theory of relativity and deductions from it cannot be applied to all bodies in nature. It holds only on the molar level for measurements of molar and microscopic phenomena which are referred to molar reference frames. It owes its existence to the epistomological circumstance that the macroscopic atom, like the microscopic atoms, is not immediately visible. This makes it impossible to determine whether a given molar object is at rest or in motion relatively to it. This means that the principle of the relativity of mass holds only when the relativity of motion holds, and that it cannot be applied in a literal metaphysical sense to the microscopic atomic entities. It must be maintained, therefore, that the doctrine of the complete annihilation of atomic matter can receive no justification from relativity physics.

Finally, the second law of thermo-dynamics cannot be valid for the whole of nature. No energy can escape from this finite universe, nor can the present state of heterogeneous chemical organization decrease to a homogeneous minimum. The physical macroscopic atom prevents both of these consequences. This confirms the conclusions of Millikan concerning the second law.

I must refer, in closing, to A. N. Whitehead. He was the first one to

discover the real difficulties to which the theory of relativity has given rise. Without his analysis, the contradictions which form the basis of this paper would not have been discovered.

[1] A. S. Eddington, *Space, Time and Gravitation*, Cambridge, **1921**, p. 197.

[2] A. Einstein, *Math. Annalen*, **97**, p. 100.

[3] A. S. Eddington, *The Mathematical Theory of Relativity*, Cambridge, **1924**, p. 213–4.

[4] Sir I. Newton, *The Mathematical Principles of Natural Philosophy*, Scholium foll. Def. 8.

[5] J. Perrin, *Atoms*, London, **1923** (83–133).

[6] J. H. Jeans, *Astronomy and Cosmogony*, Cambridge, **1928** (345).

[7] A. N. Whitehead, *The Principles of Natural Knowledge*, Cambridge, **1919**, Ch. 4. *The Principle of Relativity*, Cambridge, **1922**, Ch. 3.

[8] A. Einstein, *Sits. d. Pr. Akad. d. W.* **1917.**

[9] A. S. Eddington, *The Mathematical Theory of Relativity*, Cambridge, **1924** (162).

[10] W. Heisenberg, *Zeit. f. Physik.* **43** (879–93).

[11] F. S. C. Northrop, *J. of Phil.*, New York, 25 (449–467).

[12] H. Weyl, "Gravitation and Electricity," *Sits. d. Pr. Akad. d. W.*, 1918.

[13] F. S. C. Northrop, *The Unitary Field Theory of Einstein in Its Bearing on the Macroscopic Atomic Theory.*

[14] J. H. Jeans, *Astronomy and Cosmogony*, Cambridge, **1928** (351).

THEORETICAL INTERPRETATION OF HYPER-FINE-STRUCTURE IN SINGLY IONIZED THALLIUM, Tl II

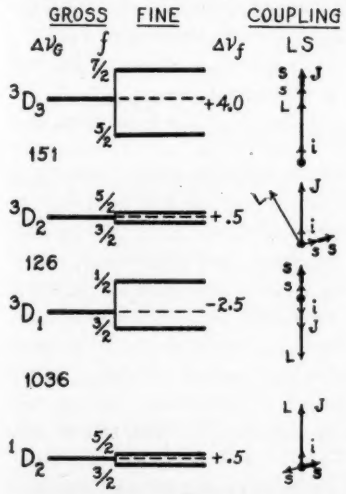
BY H. E. WHITE[1]

PHYSIKALISCH-TECHNISCHE REICHSANSTALT, CHARLOTTENBURG, GERMANY

Communicated December 7, 1929

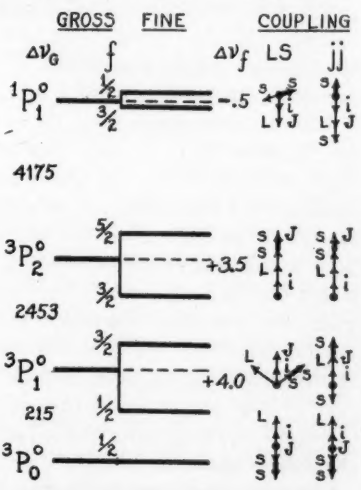
A recent extension of the multiplet structure of singly ionized thallium, made by McLennan, McLay and Crawford[2] shows that many of the terms have relatively large fine-structure separations. Since the lines, as stated by them, were photographed with an ordinary spectrograph only the widest fine-structures could be measured. The frequency number differences are therefore only given to 0.5 cm.$^{-1}$. Although this leaves much to be desired in the fine-structure measurements of Tl II, an explanation of its fine-structure is here given. This is based upon the theoretical interpretation of term fine-structures given by the author[3] for Cd I, Ba I, Tl I, Bi I, La I and La II. There it is pointed out (as is excellently confirmed in the spectra here discussed) that the widest fine-structures are found in those terms arising from electron configurations involving a single unbalanced *s* electron. An *s* electron being found part of the time near the nucleus and part of the time outside of all of the other electrons merely acts as a "means to an end" of coupling the nuclear angular momentum, **i**, with the electron resultant, **J**. Fine-structures in Bi I, La I,

Mn I, Pr II and La II show that space quantization must take place between **i** and **J**. One is therefore led to postulate that the total fine-structure separation for a single multiplet term is determined by the strength of coupling between **i** and **J**, this coupling depending upon the cos **si** and the cos **sJ**, and that the separations between fine-structure terms themselves are determined by cos **Ji** (The Landé Interval Rule). For a multiplet term arising from an electron configuration involving but one deeply penetrating *s* electron, when cos **Ji** = 1 and **J** > **i**, the fine-structure terms are (*a*) *inverted* when cos **si** is negative and (*b*) *normal* when cos **si** is positive. For nuclear angular momentum **i** due to positive charges, the words *normal* and *inverted* should be reversed in (*a*) and (*b*).



ELECTRON CONFIGURATION 6s7d

FIGURE 1



ELECTRON CONFIGURATION 6s 7p

FIGURE 2

Hyper-fine-structure of the energy levels in singly ionized thallium.

In figure 1, the fine-structures of the four levels $^3D_3$, $^3D_2$, $^3D_1$ and $^1D_2$ for the electron configuration 6*s*7*d* are shown. With $\mathbf{i} = {}^1/_2.h/_2\pi$ each level is split into two components, **i** and **J** being either in the same or in the opposite direction. The vector diagrams are in each case drawn for the term with the largest **f** value. For the $^3D_3$ term, figure 1, the *s* electron spin, $\mathbf{s} = {}^1/_2$ (the heavily drawn vector), is parallel with **J** so that when cos **Ji** = 1, cos **si** = 1 and cos **sJ** = 1; hence the large fine-structure separation. For $^3D_2$ when cos **Ji** = 1, cos **si** and cos **sJ** are very small but positive, resulting in small positive fine-structure separation. For $^3D_1$, when cos **Ji** = 1, cos **si** = −1 and cos **sJ** = −1 which results in

a quite large inverted fine-structure separation. For the $^1D_2$ term the direction of the electron spins are rather indeterminate from the multiplet structure, but from the fine-structure it can be said that when cos **Ji** = 1, cos **si** and cos **sJ** should be small positively. An analogous case of this turning over process is found in singly ionized lanthanum.[3] In lanthanum, however, each of the ^3D *6s5d* terms have more than two components.[4]

In figure 2, the $^3P^\circ_2$, $3P^\circ_1$, $3P^\circ_0$ and $^1P^\circ_1$ terms of *6s7p* are shown. The $^3P^\circ$ gross-structure separations indicate more nearly a **jj** coupling than they do an **LS**. For this reason both types of vector coupling are shown in the figure. The fine-structures of $^3P^\circ_2$ and $^3P^\circ_0$ need no explanation. A **jj** coupling for $^3P^\circ_1$ should show fairly wide normal fine-structure separation while an **LS** coupling should show somewhat smaller normal separation. The fine-structure as well as the gross-structure here favors a **jj** coupling. In the case of the $^1P_1$ term **jj** coupling should show inverted terms as observed, while **LS** coupling is indeterminate.

The $^3P^\circ_2$ *6s6p* separation, 3.5 cm.$^{-1}$, and the $^3P^\circ_2$ *6s7p* separation, 3.5 cm.$^{-1}$, show that the large fine-structure separations are due primarily to the *6s* electron, and that going to the limit in this series the $^2S_{1/2}$ *6s* term must have a separation not far from 3.5 cm.$^{-1}$ (*6s7p* is not far from the limit). Practically all of the observed terms of Tl II have $^2S_{1/2}$ *6s* as a limit and therefore in series all of the fine-structures must approach in width the $^2S_{1/2}$ separation. An analogous case to this is found in Cd I where the first three members of the series $^3S_1$ *5sms* (*m* = *6*, *7* and *8*) have separations 0.397, 0.375 and 0.364 cm.$^{-1}$, respectively. Here the limit is obviously $^2S_{1/2}$ *5s* with doublet fine-structure separation about 0.350 cm.$^{-1}$. The largest fine-structure observed by McLennan, McLay and Crawford[2] in Tl II is for the $^3S_1$ *6s7s* term (5.0 cm.$^{-1}$). The role of an *s* electron is again revealed.

Observations of fine-structure in Tl I by Schüler and Brück[5] indicates an isotope of thallium with nucleur angular momentum **i** = 0. The energy levels for such atoms, shown dotted in figures 1 and 2, lie at the center of gravity of the two fine-structure levels, the latter having the relative weights given by 2**f** + 1. Such weighing of fine-structure terms is excellently verified in the spectra of Cd I[3]. Transitions between these dotted levels, where **i** = 0, would not be observed by McLennan, McLay and Crawford since they should be rather faint lines and in most cases fall very close to the strongest fine-structure line component. From accurate measurements of fine-structure separations, the following information concerning an element can be quite definitely determined: (*a*) whether the nucleur spin, **i**, is due to positive charges or negative, (*b*) the value of **i** in units of $h/2\pi$, (*c*) whether there are isotopes with different values of **i**, and (*d*) the angle **s** (or more generally **j**), of the deep penetrating electron, makes with the **J** of each term.

The theoretical interpretation of the hyper-fine structure of singly ionized thallium, given above, does not depend upon thallium alone, but also upon analogous cases in other spectra.

[1] NATIONAL RESEARCH FELLOW.

[2] McLennan, McLay and Crawford, *Proc. Roy. Soc.* **125A**, 570, 1929.

[3] White, *Phys. Rev.*, in printer's hands, 1929.

[4] Meggers and Burns, *J. O. S. A.*, **14**, 449, 1927.

[5] Schüler and Brück, unpublished data.

THE DISTRIBUTION OF LIGHT INTENSITY IN A FRESNEL DIFFRACTION PATTERN FROM A STRAIGHT EDGE

BY THEODORE LYMAN

JEFFERSON PHYSICAL LABORATORY, HARVARD UNIVERSITY

Read before the Academy November 19, 1929

Fresnel in his celebrated memoir calculated the distribution of the intensity of the light in the diffraction pattern from a straight edge for the case where source and screen are separated by a finite distance. He then verified the results of his calculation by experiment with such a satisfactory degree of accuracy that no one since his time has been tempted to repeat his work. His quantitative results, however, were concerned only with the positions of the maxima and minima in the pattern; he did not possess the means of measuring the relative intensity of the light. As far as I am aware, all those who have studied diffraction patterns since the time of Fresnel have also confirmed their experimental investigations to determinations of the positions of maxima and minima (see, for example, Hufford and Davis, *Phys. Rev.*, **33**, p. 589, 1929).

In view of the interest that is taken nowadays in the theory of diffraction phenomena of this type it seemed worth while to make an attempt to measure the relative intensity of the light in various parts of the pattern from a straight edge.

In this attempt I have employed photographic photometry. The arrangement of the experiment is simple. Light from a vertical mercury arc was focused on the slit of a monochromatic illuminator which in turn focused an image of the line 4358 A.U. on the slit of the diffraction apparatus. This second slit, 0.02 millimeter wide and one millimeter high, was fixed in one end of a piece of common six-inch stove pipe, three meters long and lined with black velvet; the other end of the pipe carried the plate holder. The aperture at which diffraction occurred followed the original design of Fresnel. It was bounded by two parallel straight edges three centimeters high usually separated by a distance of about 1.5

cm. This aperture was mounted on a diaphragm inserted in the stove pipe at a distance of one meter from the slit; thus the distance of the photographic plate from the straight edges was two meters. The face of the photographic plate was protected by a shield which reduced the height of the image to about 1.3 cm.

This image on the photographic negative consisted of the separate diffraction patterns produced by the two straight edges acting independently; these patterns merged into a region of constant illumination at points near the axis of the stove pipe.

Considerable care was necessary in order to secure symmetry of illumination of the two straight edges. The success of the adjustment was determined by the equality of the two patterns.

The object of the experiment was to measure the distribution of opacity in the photographic negative and then to calculate from these data the ratios of the intensity of the first maximum to the intensities of the succeeding maxima and minima. In this process the determination of the characteristic curve of the photographic plate is necessary.

It may be well to remind the general reader of the method of calibrating a photographic plate. A small area was first exposed to a source of light for a given time, the intensity of the light was then cut down step by step by known amounts, at each step an exposure was made always for the same time. The resulting negative then exhibited a series of areas of varying opacity produced by light of different intensities, the ratio of these intensities being known.

The opacity, I_0/I, was measured, and the logarithm of the quantity or the density, was plotted against the logarithm of the intensity of the light expressed in per cent. The resulting curve is the characteristic for the particular plate.

The intensity of the light was reduced in these experiments in two ways. *First*, by interposing screens, composed of photographic negatives, between the mercury arc and the slit of the monochromatic illuminator. The specular transmission of these screens for $\lambda 4358$ was determined at the Bureau of Standards. *Second*, by covering a portion of the photographic negative with a set of screens; their diffuse transmission was determined here with a Bausch and Lomb polarizing photometer for white light.

In the first method the diffraction pattern was photographed, the time of exposure being of the order of an hour. A series of exposures was then made on the same plate with the screens interposed successively in front of the monochromator. A suitable motion of the photographic plate and of a diaphragm in front of it, placed these exposures side by side and close to the axis of the apparatus. To make sure that diffraction bands did not interfere with the uniformity of illumination, the distance between the straight edges was increased during these exposures. To reduce the

time the slit of the diffraction apparatus was widened. The actual time of each of these calibrating exposures was about seven minutes.

In the second method, the calibrated screens were placed nearly in contact with the plate. The diffraction pattern and the data for the characteristic curve were obtained simultaneously with an exposure of one hour.

During the progress of the investigation many photographs were made. Of these the five most satisfactory were selected for measurement.

Eastman No. 36 plates were used; they were developed by the brush method for two and a half minutes, at 65°. For Plate No. 1, Sterling M. Q. developer was employed, in the other cases Eastman special developer was used.

In Plates 1, 2 and 5 the calibrating screens were placed in front of the slit of the monochromatic illuminator. In 3 and 4 they were placed in front of the photographic plate.

For the first four plates the distance between the straight edges was 1.5 centimeters; in the case of No. 5 the distance was 1 centimeter.

In plates 1, 3 and 5 the symmetry of the diffraction pattern was excellent, that is to say, the opacity at the first maximum produced by either straight edge differed from the mean by about a half per cent of the mean. In cases 2 and 4 the symmetry was only fair; the difference being about three per cent of the mean.

The time of exposure in case 1 was 30 minutes, in case 2, 80 minutes, and in the remaining cases one hour.

The distribution of the opacity in the diffraction pattern and the opacity of the calibration blocks were measured on the recording microphotometer at this laboratory. The opacities for each of the two patterns were determined and the mean for the homologous values was taken.

The characteristic curves over the range of densities covered by the diffraction patterns were straight lines in cases 3, 4 and 5. In those cases where the screens were placed over the photographic plate a slight extrapolation of the characteristic was necessary. The patterns in Nos. 1 and 2 were somewhat underexposed with the result that the lower densities in the two cases fell on the curved part of the characteristic.

In considering the sources of error which effect the experimental results it is to be remembered that where the screens, calibrated for a definite wave-length by the Bureau of Standards, were used the data are probably more trustworthy than in the other cases. The errors introduced by uncertainties in the readings of the microphotometer are relatively small. The chief source of trouble lies in the lack of uniformity in the behavior of the photographic plate and the consequent uncertainty as to the exact position of the calibration curve. Furthermore, the assumption that the fog correction for the pattern was the same as that for the curve may not be entirely justified.

All things considered I estimate that the experimental results may be in error by four per cent.

The table exhibits the results of the investigation. In it the ratio of the first maximum to the succeeding minima and maxima as computed by the method of Fresnel (Mascart, *Traité d'Optique*, Vol. I, p. 280) are compared with the same ratios as determined by experiment.

The agreement between theory and experiment appears to be within the limit of error.

TABLE 1

	EXPERIMENT					
THEORY	1	2	3	4	5	THEORY
1.762	1.76	1.82	1.70	1.70	1.78	1.762
1.145	1.16	1.16	1.16	1.15	1.14	1.145
1.625	1.61	1.68	1.59	1.57	1.62	1.625
1.191	1.21	1.22	1.21	1.19	1.16	1.191
1.572	1.55	1.59	1.53	1.53	1.57	1.572

I am indebted to my colleague, Dr. Crawford, for the microphotometer records and to my assistant, Mr. H. W. Leighton, for skillful experimental manipulation.

NOTE ON RANGE-FINDING

BY FRANK MORLEY AND HENRY A. ROBINSON

THE JOHNS HOPKINS UNIVERSITY

Communicated December 12, 1929

1. *The Problem on a Plane.*—The problem is the location of an object such as a gun P by timing a sound or other vibration of known velocity at different stations P_i $(i = 1, 2, \ldots, n)$. The time for a fixed longitude recorded at P_i multiplied by the known velocity is a known distance d_i. Denote the distance between P and P_i by r_i. Then for any two stations P_1 and P_2, we have

$$r_1 - r_2 = d_1 - d_2,$$

so that

$$r_i = d_i + k \quad (i = 1, 2, \ldots, n). \tag{1}$$

For three stations we have, then, in determinant form,

$$\left| r_i^2 - d_i^2,\ d_i,\ 1 \right| = 0. \tag{2}$$

This is the equation of a line on the gun.

If we were restricted to three stations, the problem would be to find the centers of the two circles which touch properly three given directed

circles. The three circles here are the circles (P_i, d_i) and from (1) k is the radius of the touching circle. To identify the line, we notice that (2) is true when

$$r_1^2 - d_1^2 = r_2^2 - d_2^2 = r_3^2 - d_3^2,$$

that is, it is true for the radical center. Again it is true when

$$(r_1^2 - d_1^2)/d_1 = (r_2^2 - d_2^2)/d_2 = (r_3^2 - d_3^2)/d_3. \qquad (3)$$

For two directed circles (P_1, d_1) and (P_2, d_2) there is one system of properly tangent circles, all being orthogonal to the circle

$$(r_1^2 - d_1^2)/d_1 = (r_2^2 - d_2^2)/d_2.$$

This then is the circle of an inversion which sends each tangent circle into itself. For the three circles (3) there are three such circles, and the line (2) is their radical axis. It is the line discussed in article 118 of Salmon's *Conic Sections*.

2. *Coördinates of the Gun.*—It is proper to take four stations for the planar problem. Using rectangular coördinates, let P_i be the point (x_i, y_i). Then for three stations, the line (2) is

$$\left| (x - x_i)^2 + (y - y_i)^2 - d_i^2, d_i, 1 \right| = 0$$

or

$$2x \left| x_i, d_i, 1 \right| + 2y \left| y_i, d_i, 1 \right| = \left| x_i^2 + y_i^2 - d_i^2, d_i, 1 \right|.$$

For four stations we have four such equations. To eliminate y we multiply each by the omitted y_i and so build

$$\left| y_i, y_i, d_i, 1 \right|$$

which is zero. We have then

$$2x \left| x_i, y_i, d_i, 1 \right| = \left| x_i^2 + y_i^2 - d_i^2, y_i, d_i, 1 \right|, \qquad (4)$$

giving the x-coördinate of the gun. For the y-coördinate we interchange x and y.

3. *Elimination of Temperature.*—The velocity of sound is a function of the temperature. If we assume that the function is linear, it is proper to take five stations. We have to get rid of

$$\left| d_i^2, y_i, d_i, 1 \right|,$$

so that from the five equations (4) we build

$$\left| d_i^2, d_i^2, y_i, d_i, 1 \right|$$

which is zero. We have then

$$2x \left| x_i, y_i, d_i^2, d_i, 1 \right| = \left| x_i^2 + y_i^2, y_i, d_i^2, d_i, 1 \right|, \qquad (5)$$

which being homogeneous in the d_i needs no correction for the temperature.

4. *The Problem on a Space.*—For three stations on a space (3) is

$$2x\,|\,x_i, d_i, 1\,| + 2y\,|\,y_i, d_i, 1\,| + 2z\,|\,z_i, d_i, 1\,| = |\,x_i^2 + y_i^2 + z_i^2 - d_i^2, d_i, 1\,|.$$

For four stations on a space we eliminate y, as above, and have

$$2x\,|\,x_i, y_i, d_i, 1\,| + 2z\,|\,z_i, y_i, d_i, 1\,| = |\,z_i^2 + y_i^2 + z_i^2 - d_i^2, y_i, d_i^2, 1\,|.$$

For five stations on a space we have five such equations. Multiplying by the omitted z_i and forming

$$|\,z_i, z_i, y_i, d_i, 1\,|,$$

we have

$$2x\,|\,x_i, y_i, z_i, d_i, 1\,| = |\,x_i^2 + y_i^2 + z_i^2 - d_i^2, y_i^2, z_i^2, d_i^2, 1\,|, \qquad (6)$$

giving a typical coördinate of the gun. As in Article 3, this may be made independent of the temperature by the use of six stations.

5. *The Problem on a Sphere.*—When the vibration travels along the surface of a sphere (as in a tidal wave to a first approximation) the equations (1) refer to the lengths of arcs of great circles. We have then

$$\cos r_i = \cos d_i \cos k - \sin d_i \sin k. \qquad (7)$$

The cosine is necessary, because for constants m_i a circle on the sphere is given by

$$\Sigma m_i \cos r_i = m_0.$$

In Salmon's *Geometry of Three Dimensions*, chapter X, the sine of the arcual distance from a variable point to a great circle is employed. But here it is more convenient to think of this as the cosine of the arc from a variable to a fixed point. We have from (7) for three stations

$$|\,\cos r_i, \cos d_i, \sin d_i\,| = 0, \qquad (8)$$

a great circle on the source of the wave.

It is worth while to mention that the circles of inversion (3) of the planar case are here replaced by

$$(\cos r_i - \cos d_i)/\sin d_i = (\cos r_2 - \cos d_2)/\sin d_2 = (\cos r_3 - \cos d_3)/\sin d_3.$$

Let now the longitude and colatitude of the point P_i be a_i and b_i, those of the source a and b. Then

$$\cos r_i = \cos b \cos b_i + \sin b \sin b_i \cos (a - a_i),$$

so that (8) becomes

$$\begin{aligned} &\cos b\,|\,\cos b_i, \cos d_i, \sin d_i\,| \\ &\quad + \sin b \cos a\,|\,\sin b_i \cos a_i, \cos d_i, \sin d_i\,| \\ &\quad + \sin b \sin a\,|\,\sin b_i \sin a_i, \cos d_i, \sin a_i\,| = 0. \end{aligned}$$

We have then for four observations

$$\cot a \,|\cos b_i,\ \sin b_i \cos a_i,\ \cos d_i,\ \sin d_i| + |\cos b_i,\ \sin b_i \sin a_i,\ \cos d_i,\ \sin d_i| = 0, \tag{9}$$

and

$$\cot b \,|\cos b_i,\ \sin b_i \cos a_i,\ \cos d_i,\ \sin d_i| + \cos a \,|\sin b_i \cos a_i,\ \sin b_i \sin a_i,\ \cos d_i,\ \sin d_i| = 0, \tag{10}$$

giving the longitude and the colatitude of the source.

APPLICATIONS OF AN EXPANSION THEOREM TO THE DEVELOPMENT OF THE DISTURBING FUNCTION

BY ERNEST W. BROWN

DEPARTMENTS OF MATHEMATICS AND ASTRONOMY, YALE UNIVERSITY

Communicated November 29, 1929

1. The simplest form of the theorem is stated as follows. If $f(a + \alpha p)$ be a function developable in powers of α and if $D \equiv \alpha\partial/\partial\alpha$, then

$$f(a + \alpha p) = p^D f(a + \alpha). \tag{1.1}$$

It is assumed that p^D is developed in integral powers of D and, after development, is applied to $f(a + \alpha)$.

The proof is simple. If n be an integer,

$$D^n \alpha^m = m^n \alpha^m.$$

Hence, if $\Psi(D)$ be a function of D developable in integral powers of D,

$$\Psi(D)\alpha^m = \Psi(m)\alpha^m. \tag{1.2}$$

Suppose $\Psi(m) = p^m$. Then

$$p^m \alpha^m = p^D \alpha^m. \tag{1.3}$$

Hence, by Taylor's theorem,

$$f(a + \alpha p) = \Sigma_m \frac{p^m \alpha^m}{m!} f^{(m)}(a) = p^D \frac{\Sigma \alpha^m}{m!} f^{(m)}(a) = p^D f(a + \alpha).$$

2. Examples. Let

$$p = (1 + q)^j,$$

where $|q| < 1$ and j is any quantity. Then

$$p^D = (1 + q)^{jD} = 1 + jD.q + \frac{jD(jD - 1)}{1.2} q^2 + \ldots$$

Hence the coefficient of q^n in the expansion of

$$f\{a + \alpha(1 + q)^j\}$$

in powers of q is

$$\frac{jD(jD - 1) \ldots (jD - n + 1)}{1.2 \ldots n} f(a + \alpha).$$

This example illustrates the manner in which the theorem is applied to the development of the disturbing function.

As a second example, let $p = e^Z$, so that

$$p^D = e^{ZD} = 1 + ZD + \frac{Z^2D^2}{1.2} + \ldots,$$

and the coefficient of Z^n in the expansion of

$$f(a + \alpha e^Z)$$

in powers of Z is

$$\frac{D^n}{n!} f(a + \alpha).$$

Many expansions which are usually quite troublesome are made with comparative ease by the theorem.

3. In the development of the disturbing function we have in the first place to expand functions of the form $(1 - 2\alpha \cos A + \alpha^2)^{-s}$ as Fourier series,

$$a_s^{(0)} + 2\Sigma_i a_s^{(i)} \cos iA, \qquad i = 1, 2, \ldots$$

where $|\alpha| < 1$ and s is an odd multiple of $1/2$; the latter limitation is not necessary for the application of the theorem. The values of the coefficients for any given numerical value of α are obtained by well-known methods. Further, equally well-known methods enable us to obtain the derivatives of these coefficients with respect to α or $\log \alpha$.

If we put

$$\text{Exp. } A\sqrt{-1} \equiv y,$$

so that

$$1 - 2\alpha \cos A + \alpha^2 = (1 - \alpha y)(1 - \alpha/y),$$

we can state that $a_s^{(i)}$ is the coefficient of y^i or of y^{-i} in the expansion of

$$(1 - \alpha y)^{-s} (1 - \alpha/y)^{-s}$$

in positive and negative powers of y. This coefficient is

$$a_s^{(i)} = \Sigma_j \binom{-s}{i+j}\binom{-s}{j} \alpha^{i+2j},$$

in which the brackets denote binomial coefficients.

4. Consider next the expansion of

$$(1 - p\alpha y)^{-s} (1 - q\alpha/y)^{-s}$$

The coefficient of y^i (i positive) is obtained by finding the product of the coefficient of y^{i+j} in the expansion of the first factor, with the coefficient of y^{-j} in the expansion of the second factor, and summing for all positive values of j. With the usual notation for binomial coefficients this sum is

$$(-1)^i \Sigma_j \binom{-s}{i+j} \binom{-s}{j} p^{i+j} q^j \alpha^{i+2j}.$$

The theorem is applied by putting in (1.2)

$$m = i + 2j \quad , \quad \Psi(m) = p^{i/2+m/2} q^{-i/2+m/2}.$$

It gives

$$p^{i+j} q^i \alpha^{i+2j} = p^{D/2+m/2} q^{D/2-m/2} a^{i+2j}.$$

The coefficient of y^i may therefore be written

$$(-1)^i p^{D/2+m/2} q^{D/2-m/2} \Sigma_i \binom{-s}{i+j} \binom{-s}{j} \alpha^{i+2j} = p^{D/2+m/2} q^{D/2-m/2} a_s^{(i)}.$$

The coefficient of y^{-i} is evidently got by interchanging p, q, that is, by changing the sign of i while remembering that $a_s^{(i)}$ is unaltered by the change.

5. The result of § 4 is particularly useful whenever p, q are products of binomial factors of the form

$$p = (1 - eZ)^a (1 - e/Z)^b \quad , \quad q = (1 - eZ)^c (1 - e/Z)^d, \quad (5.1)$$

because the indices of $1 - eZ$, $1 - e/Z$ are then, respectively,

$$\lambda \equiv \tfrac{1}{2}(a + c)D + \tfrac{1}{2}(a - c)i \quad , \quad \mu \equiv \tfrac{1}{2}(b + d)D + \tfrac{1}{2}(b - d)i.$$

The coefficient of $y^i e^{j+2j_1} Z^j$ is

$$(-1)^j \binom{\lambda}{j+j_1} \binom{\mu}{j_1} a_s^{(i)}.$$

This satisfies the condition of being expansible in powers of D since D is only present in the form of a product of $j + 2j$, factors each of the form $fD + g$.

If p, q contain also powers of two other factors $1 - e'Z'$, $1 - e'/Z'$ of the same form as (5.1) with all letters in the right hand members having accents, we obtain for the coefficient of $y^i e^{j+2j_1} e'^{j_1+2j_1'} Z^j Z'^{j'}$,

$$(-1)^{j+j'} \binom{\lambda}{j+j_1} \binom{\mu}{j_1} \binom{\lambda'}{j'+j_1'} \binom{\mu'}{j_1'} \quad ,$$

where

$$\lambda' \equiv \tfrac{1}{2}(a' + c')D + \tfrac{1}{2}(a' - c')i \, , \, \mu' = \tfrac{1}{2}(b' + d')D + \tfrac{1}{2}(b' - d')i.$$

The coefficients of negative powers of Z, Z' are obtained by obvious interchanges of symbols.

Evidently the result can be extended to any number of such pairs of factors; in the development of the disturbing function only two such pairs are needed. The importance of the result arises from the fact that *the coefficient of any power of y, e, e', Z, Z' can be at once written down as a product of linear operators of the form fD + g, these products being derived directly from the coefficients in certain binomial expansions.* The number of such linear operators is the sum of the exponents of e, e'. The final step consists in multiplication of the operators so as to arrange them in powers in D which then act on $a_z^{(i)}$.

6. I shall only indicate here the manner in which these formulæ may be applied to the expansion of the disturbing function, reserving the detailed work for publication elsewhere.

Let r be the radius vector of a point on an ellipse with the origin in one focus. Let $\tilde{\omega}$ be the longitude of the nearer apse, f the true anomaly, X the eccentric anomaly, M the mean anomaly, $2a$ the major axis, e the eccentricity. Let accented letters denote the corresponding quantities for a point on another ellipse with the same origin. When the ellipses are in the same plane, the square of the distance between the two points is

$$r'^2 + r^2 - 2rr' \cos (f + \tilde{\omega} - f' - \tilde{\omega}').$$

The disturbing function is the inverse of the square root of this expression.

When the ellipses are not in the same plane, it can be shown that the expansion of the disturbing function depends mainly on the expansion of expressions of the form

$$\left\{1 + \frac{r^2}{r'^2} - 2\frac{r}{r'} \cos (f + \tilde{\omega} - f' - \tilde{\omega}')\right\}^{-s},$$

where s is an odd multiple of $1/2$. If we put

$$u \equiv \text{exp.}\ (f + \tilde{\omega} - f' - \tilde{\omega}')\sqrt{-1},$$

this expression can be written

$$\left(1 - \frac{r}{r'}u\right)^{-s}\left(1 - \frac{r}{r'}\frac{1}{u}\right)^{-s}.$$

The expansion takes different forms according as we expand in multiples of the true, eccentric or mean anomalies.

7. *Expansion on Multiples of the True Anomalies.* We have

$$\frac{a}{r} = \frac{1 + e\cos f}{1 - e^2}\ ,\quad \frac{a'}{r'} = \frac{1 + e'\cos f'}{1 - e'^2}.$$

If we put

$$e \equiv \frac{2\eta}{1 + \eta^2}\ ,\quad \phi \equiv \text{exp.}\ f\sqrt{-1},$$

we obtain

$$\frac{a}{r} = \frac{1+\eta^2}{(1-\eta^2)^2}(1+\eta\phi)(1+\eta/\phi).$$

With a similar formula for r', it is evident that we can apply the formulæ of § 5 if we put

$$y = u \quad , \quad \alpha = \frac{a}{a'}\frac{(1-\eta^2)^2}{(1-\eta'^2)^2}\frac{1+\eta'^2}{1+\eta^2},$$

$$p = q = (1+\eta\phi)^{-1}(1+\eta/\phi)^{-1}(1+\eta'\phi')\ (1+\eta'/\phi').$$

We then obtain the required expansion in powers of η, η', ϕ, $1/\phi$, ϕ', $1/\phi'$. The trigonometric forms are as usual obtained by putting $\phi^i = \cos if + \sqrt{-1}\sin if$, etc.

8. *Expansion in Multiples of the Eccentric Anomalies.* We have the relations

$$r\cos f = a(\cos X - e),\ r\sin f = a(1-e^2)^{1/2}\sin X,$$

If we put

$$\chi \equiv \exp. X\sqrt{-1} \quad , \quad e \equiv \frac{2\eta}{1+\eta^2},$$

with the same meaning as before for ϕ, the relations can be written

$$(1+\eta^2)r\phi = a\chi\left(1-\frac{\eta}{\chi}\right)^2 \quad , \quad (1+\eta^2)\frac{r}{\phi} = \frac{a}{\chi}(1-\eta\chi)^2,$$

with similar expressions for r', ϕ'.

In this case the formulæ of §5 are applied by putting

$$\alpha = \frac{1+\eta'^2}{1+\eta^2}\frac{a}{a'} \quad , \quad y = \exp.(\chi + \bar{\omega} - \chi' - \bar{\omega}')\sqrt{-1},$$

$$p = \left(1-\frac{\eta}{\chi}\right)^2\left(1-\frac{\eta'}{\chi'}\right)^{-2} \quad , \quad q = (1-\eta\chi)^2(1-\eta'\chi')^{-2}.$$

The remaining steps are the same as in §7.

9. *Expansion in Multiples of the Mean Anomalies.* The formulæ for the true and eccentric anomalies are exact as shown in the two previous paragraphs. Those given here for the mean anomalies are necessarily approximate only since the coördinates cannot be expressed in terms of them except by the use of infinite series. The main point is, however, that these approximate formulæ contain the terms which cause slow convergence, and that if the remainders were neglected the resulting errors would be small. These remainders are, however, of the same order as the terms depending on the inclination and the position of the node and can be treated with them by the same methods. Thus while the formulæ do

not present the same simplicity as those for the previous cases they are nearly as accurate and as effective in abbreviating the work in actual calculation, except perhaps where very low accuracy is required. I give here only the main outline, reserving for publication elsewhere the details necessary for practical application.

If l be the mean anomaly and $\lambda \equiv \exp. l\sqrt{-1}$, we have from Cayley's tables,[1] as far as e^3,

$$\frac{r}{a}\frac{\phi}{\lambda} = 1 - \frac{1}{2}e^2 + \left(\frac{1}{2}e - \frac{3}{8}e^3\right)\lambda - \frac{3}{2}\frac{e}{\lambda}$$
$$+ \frac{3}{8}e^2\lambda^2 + \frac{1}{8}\frac{e^2}{\lambda^2} + \frac{1}{3}e^3\lambda^3 + \frac{1}{24}\frac{e^3}{\lambda^3} = \left(1 + \frac{1}{4}e^2\right)\left(1 - \frac{e}{\lambda}\right)^{3/2}(1 - e\lambda)^{-1/2}$$
$$- \frac{1}{4}\frac{e^2}{\lambda^2} + \frac{e^3}{16}\left(\lambda + \frac{1}{3}\lambda^3 + \frac{3}{\lambda} - \frac{1}{3}\frac{1}{\lambda^3}\right).$$

Thus the error caused by using the finite expression is of order $e^2/4$. Further, the continuation of the expansion to higher powers of e, shows that the convergence of the remainder is considerably more rapid than that of the series for $r\phi$.

For the finite expression we put

$$\alpha = \frac{1 + {}^1/_4e^2}{1 + {}^1/_4e'^2}\frac{a}{a'}, \quad q = \exp.\,(l + \tilde{\omega} - l' - \tilde{\omega}')\sqrt{-1},$$
$$p = \left(1 - \frac{e}{\lambda}\right)^{3/2}(1 - e\lambda)^{-1/2}\left(1 - \frac{e'}{\lambda'}\right)^{-3/2}(1 - e'\lambda')^{1/2},$$
$$q = \left(1 - \frac{e}{\lambda}\right)^{-1/2}(1 - e\lambda)^{3/2}\left(1 - \frac{e'}{\lambda'}\right)^{1/2}(1 - e'\lambda')^{-3/2}.$$

10. The theorem of § 1 substantially gives a simple method for the formal expansion of a function of a function. It can be extended to a function of n functions, each of which is a function of m variables. I am, however, concerned here less with its more general aspects than with its use in adapting expansions which are known to be theoretically possible to forms convenient for calculation. This latter, in the present day, is in fact the fundamental problem of applied mathematics, and more particularly of celestial mechanics. In the planetary theory, a great part of the labor consists in getting a very few—often only two or three—coefficients in the expansion of the disturbing function to a high degree of accuracy, the remaining coefficients being needed to a much lower degree of accuracy. It is in the solution of this problem that the results of this paper appear to be particularly useful.

The abbreviation of the labor is most notable in the case of expansion in terms of the eccentric anomalies, the detailed work for which I have given in pages 19–23 of No. 933 of the *Astronomical Journal*. Newcomb's

expansion in terms of these variables requires much space in order to set down the formulæ for the operators to the sixth powers of the eccentricities. By the formulæ of this paper these operators are at once written down to any order whatever from a general formula which in form is nothing else than a combination of the coefficients in the binomial theorem.

[1] *Mem. Roy. Astron. Soc.*, **29**, pp. 260, 261 (1860).

[2] *Astron. Pap. Amer. Eph.*, **3**, pp. 1–200 (1891).

ON CERTAIN POLYNOMIAL AND OTHER APPROXIMATIONS TO ANALYTIC FUNCTIONS

BY HILLEL PORTISKY*

GENERAL ELECTRIC COMPANY, SCHENECTADY, N. Y.

Communicated December 10, 1929

The proof of the following results has been given and will appear in amplified form elsewhere, along with certain related results, at present obtained.

Consider first an analytic function of the complex variable z, $f(z)$, and two given points $z = a$, $z = b$, $a \neq b$. A unique polynomial $P_{2,n}(z)$ of degree at most equal to $2n - 1$ exists, whose derivatives of orders 0, 2, ..., $2n - 2$ at $z = a$, $z = b$ are equal, respectively, to the corresponding derivatives of $f(z)$ at a and b (by a derivative of order 0 is understood the function itself). Sufficient conditions for the convergence of $P_{2,n}(z)$ to $f(z)$ as n becomes infinite, are given by Theorem A; Theorem B gives necessary and sufficient conditions for the convergence of $P_{2,n}(z)$ (but not necessarily to $f(z)$); Theorem C then gives necessary conditions for the convergence of $P_{2,n}(z)$ to $f(z)$, that are almost sufficient.

We refer to a function $g(z)$ as *even about a point* $z = c$ if $g(c + z) = g(c - z)$ for all z; a similar meaning will be attached to the term *odd about a point.*

THEOREM A. *Let*

$$f(z) = e(z) + o(z),$$

where $e(z)$ is even about $(a + b)/2$, and $o(z)$ odd about $(a + b)/2$. In order that $P_{2,n}$ (z) converge to $f(z)$ for all z, and uniformly in any finite region, it is sufficient that $f(z)$ be an integral (entire) function of z, and that

$$e(z) = O(e^{k|z|}) \quad , \quad o(z) = O(e^{k'|z|}) \quad , \quad k < \pi/|a - b| \quad , \quad k' < 2\pi/|a - b|,$$

where k, k' are constants. The constants $\pi/|a - b|$, $2\pi/|a - b|$ above cannot be replaced by any larger values.

THEOREM B. *In order that $P_{2,n}(z)$ converge for all z, it is necessary and sufficient that each of the two sums*

$$\sum_{n=0}^{\infty} (-1)^n [f^{(2n)}(a) + f^{(2n)}(b)][\pi/(a-b)]^{2n},$$

$$\sum_{n=0}^{\infty} (-1)^n [f^{(2n)}(a) - f^{(2n)}(b)][2\pi/(a-b)]^{2n} \quad (1)$$

converge. In the latter case the convergence of $P_{2,n}(z)$ is uniform over any finite region, and if

$$l(z) = l_e(z) + l_0(z),$$

where $l(z)$ is the limit function, and where $l_e(z)$ is even about $(a+b)/2$, while $l_0(z)$ is odd about $(a+b)/2$, then $l(z)$ is an integral function, and

$$l_e(z) = O(e^{k|z|})$$

for any k greater than $\pi/|a-b|$, while

$$l_o(z) = O(e^{k'|z|})$$

for any k' greater than $2\pi/|a-b|$.

THEOREM C. *In order that $P_{2,n}(z)$ converge to $f(z)$ for all z, it is necessary that the two sums (1) converge, and that*

$$e(z) = O(e^{k|z|}),$$
$$o(z) = O(e^{k'|z|})$$

for any k, k' greater than $\pi/|a-b|$, $2\pi/|a-b|$, respectively. If these conditions are fulfilled, and $l(z)$ is the limit function, then

$$l(z) = f(z) + D \sin \pi \frac{z-a}{a-b} + D' \sin 2\pi \frac{z-a}{a-b},$$

where D, D' are constants.

Next consider, more generally, m given points, $z = a_1, a_2, \ldots, a_m$, no two of which are alike. There exists a unique polynomial $P_{m,n}(z)$ of degree at most equal to $mn-1$, whose derivatives at these m points of orders $0, m, \ldots, m(n-1)$ are equal, respectively, to those of $f(z)$ at these points. For $m = 1$, $P_{m,n}(z)$ reduces to the first n terms of the Taylor expansion of $f(z)$ about $z = a_1$; for $m = 2$, $P_{m,n}(z)$ reduces to the polynomial just considered (with $a = a_1$, $b = a_2$). For any $m > 1$ sufficient conditions for the convergence of $P_{m,n}(z)$ to $f(z)$ as n becomes infinite are given by

THEOREM D. *For $m > 1$ there exists a countable set of values of λ for which the system*

$$d^m u(z)/dz^m - \lambda^m u(z) = 0, \quad u(a_i) = 0, \quad i = 1, 2, \ldots, m,$$

possesses a non-trivial solution. This set has ∞ as its only limit point, and does not include 0. Let ρ_m be the minimum distance from $\lambda = 0$ to any of the points of this set.

In order that $P_{m,n}(z)$ converge to $f(z)$ for all z, and uniformly in any finite region, it is sufficient that

$$f(z) = O(e^{k|z|}) \quad , \quad k < \rho_m.$$

For $m = 2$, $\rho_m = \pi/|a - b|$.

The scope of this theorem for $m = 2$ is less than that of Theorem A to the extent that functions $f(z)$ for which

$$o(z) = O(e^{k|z|}), \quad k < 2\pi/|a - b|; \quad o(z) \neq O(e^{k'|z|}), \quad k' < \pi/|a-b|,$$

are included in Theorem A but not in Theorem D.

Finally consider an analytic function f of k variables, $x_1, x_2, \ldots, x_k$, where the latter range over a real, finite, simply connected, k-dimensional region R, bounded by a $(k - 1)$-dimensional variety S of sufficient regularity to insure the existence of the solution of the Dirichlet problem for R and its expression by means of Green's function. There will exist a unique function p_n such that

$$\nabla^{2n+2} p_n = 0 \quad \text{in } R,$$
$$p_n = f, \quad \nabla^2 p_n = \nabla^2 f, \quad \ldots \quad \nabla^{2n} p_n = \nabla^{2n} f \quad \text{on } S.$$

For $k = 1$ these functions reduce to the polynomials $P_{2,n}(z)$ with the range of z restricted to a real interval. Sufficient conditions for the convergence of p_n to f are given by

THEOREM E. *Let ρ be the smallest value of λ for which a non-trivial solution of*

$$\nabla^2 u + \lambda^2 u = 0 \quad \text{in } R,$$
$$u = 0 \quad \text{on } S,$$

exists. In order that p_n converge to f uniformly inside R, it is sufficient that the Taylor series of f be dominated by the Taylor series of an exponential

$$Ce^{\alpha_1 x_1 + \alpha_2 x_2 + \ldots \alpha_k x_k},$$

where C, α_i are constants, and

$$\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_k^2 < \rho^2.$$

* NATIONAL RESEARCH FELLOW in Mathematics, 1927–1929.

GROUPS WHICH ADMIT TWO-THIRDS AUTOMORPHISMS

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated November 29, 1929

A group is said to admit a two-thirds automorphism if it admits an automorphism in which exactly two-thirds of its operators correspond to their inverses. If G represents a group which admits at least one such automorphism, then G must be non-abelian for it is known that if more than half the operators of an abelian group correspond to their inverses in an automorphism of this group then all the operators must correspond to their inverses in this automorphism. It is also known that at most three-fourths of the operators of a non-abelian group can correspond to their inverses in an automorphism of this group and various properties of all the groups which admit three-fourths automorphisms have recently been determined.[1] In particular, it was proved that if a group admits one three-fourths automorphism it must admit exactly three such automorphisms. In the present article we shall prove, among other things, that a group which admits one two-thirds automorphism admits exactly four such automorphisms.

Let H be one of the largest subgroups of G which have the property that each one of their operators corresponds to its inverse in a two-thirds automorphism of G. We shall first assume that H is of index 2 under G. Since the product of two operators which correspond to their inverses in an automorphism of G corresponds to its inverse in the same automorphism if and only if these operators are commutative, it results that if s is an operator of G which corresponds to its inverse in the said two-thirds automorphism of G but is not found in H, then s is commutative with exactly one-third of the operators of H. That is, the abelian group H must admit an automorphism of order 2 in which exactly one-third of its operators correspond to themselves. This automorphism implies that H involves an operator of order 3 which is not found among those operators which correspond to themselves and hence we have incidentally proved the following theorem: *If an abelian group admits an automorphism of order 2 in which exactly one-third of the operators correspond to themselves, then it is the direct product of a group of order 3 and another group, and the operators of this group of order 3 correspond to their inverses in this automorphism.*

From what precedes it results that when H is of index 2 under G then the central quotient group of G is the symmetric group of order 6, and hence G admits exactly three two-thirds automorphism in which the largest subgroup whose operators correspond to their inverses therein is

of index 2. Such a group admits also a two-thirds automorphism in which the largest subgroup in which all the operators correspond to their inverses is of index 3. This two-thirds automorphism is characteristic while the other three are conjugate under the group of inner automorphisms of G. It has therefore been proved that whenever a group admits a two-thirds automorphism in which the largest subgroup composed of operators which correspond to their inverses under this automorphism is of index 2 it must also admit such an automorphism in which the largest such subgroup is of index 3. In particular, we have established the following theorem: *if a group admits at least one two-thirds automorphism in which the largest subgroup composed of all the operators which correspond to their inverses is of index* 2, *then it admits exactly three such automorphisms and these automorphisms are conjugate under the group.*

If a group admits a two-thirds automorphism then the index of the largest subgroup composed of operators which correspond to their inverses under this automorphism cannot exceed 3 since the group can be represented in a rectangular form with respect to this subgroup and not more than one-half of the operators of any other of these lines could correspond to their inverses under this automorphism. Moreover, when this subgroup is of index 3 it cannot be invariant since an automorphism of an abelian group in which half the operators correspond to themselves must be of order 2 while the quotient group would be of order 3. Hence it results that when H is of index 3 under G it must be transformed under G according to the symmetric group of order 6 since a non-invariant subgroup of index 3 is always transformed under the group according to this symmetric group.

The product of two operators which appear in two of the conjugates of this H but are not common thereto must be commutative with the cross-cut of these H's and must generate with this cross-cut an abelian subgroup of index 2 under G. Hence each of these operators transforms this subgroup according to an automorphism of order 2 and is commutative with exactly one-third of the operators of this subgroup. This invariant abelian subgroup is therefore the direct product of the group of order 3 and an abelian group. That is, *every group which admits an automorphism in which exactly two-thirds of the operators correspond to their inverses must involve a subgroup of index* 2 *which is the direct product of the group of order* 3 *and another abelian group.*

From this theorem it follows directly that if a group admits at least one two-thirds automorphism it must admit exactly four such automorphisms. The three of these which are conjugate must always be of order 2 but the fourth may be either of order 2 or the identity. A necessary and sufficient condition that it is the identity is that each of the three abelian subgroups of index 3 contained in G involves only operators of

order 2 besides the identity. When this automorphism is of order 2 then the four two-thirds automorphisms of G must generate the dihedral group of order 12, and the three conjugate two-thirds automorphisms of G must always generate the symmetric group of order 6. It may also be noted that *whenever at least one of the four two-thirds automorphisms of G is an inner automorphism then all of them are inner automorphisms and they generate a group of order* 6. The characteristic two-thirds automorphism must be the identity automorphism in this case.

A necessary and sufficient condition that there is at least one group of a given order g which admits two-thirds automorphisms is that g is divisible by 6, and the only case when there is only one such group is when g is of the form $6m$, where m is the product of distinct odd prime numbers. In view of the elementary properties of the category of groups which are characterized by the fact that each of them admits at least one two-thirds automorphism it is easy to determine the number of these groups of any order which is a multiple of 6. It is only necessary to observe that these groups are completely characterized by the abelian subgroups of index 3 involved therein and their cross-cut. That is, *the number of the distinct groups of order* $6m$ *which admit separately a two-thirds automorphism is equal to the sum of the numbers of the sets of subgroups of index* 2 *contained in the abelian groups of order* $2m$ *such that each set is composed of all these subgroups which are conjugate under the groups of isomorphisms of the corresponding groups of order* $2m$. For instance, when $m = 4$ there are four such groups since each of two of the abelian groups of order 8 involves only one set of conjugate subgroups of order 4 while the third involves two such sets.

[1] G. A. Miller, these PROCEEDINGS, **15**, 369, 1929.

DIFFERENTIAL GEOMETRIES OF FUNCTION SPACE

BY ARISTOTLE D. MICHAL

DEPARTMENT OF MATHEMATICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated November 16, 1929

In a previous paper[1] the author initiated the study of a species of functional differential geometries. These geometries are the function space analogues of the n-dimensional theories of affinely connected manifolds. An attempt to develop a projective theory in function space was instrumental in showing that the functional geometries which were developed in the cited paper[1] were a truncated form of a much more general situation. It is the object of the present note to set forth in outline some of the results obtained in these more general function space differential geom-

etries. A detailed treatment of the subject with proofs will appear elsewhere.

Let $\bar{y}^x$ stand for an arbitrary continuous function of x defined over the interval $a \leq x \leq b$. We shall be concerned with $1-1$ continuous functional transformations

$$y^x = y^x[\bar{y}^x] \tag{1}$$

that transform a function $\bar{y}^x$ into a continuous function y^x and that have differentials of the form

$$\delta y^x = {}_{x,}y^x\,\delta\bar{y}^{(x)} + y^x_{,t}\,\delta\,\bar{y}^t \qquad ({}_{x,}y^{(x)} \neq 0) \tag{2}$$

with non-vanishing Fredholm determinants

$$D[y^x_{,t}/{}_{t,}y^{(t)}] \neq 0. \tag{3}$$

The notation used in the above formulæ is essentially in accord with the one adopted by me in my former paper.[1] *We shall adhere to the convention of letting the repetition of a continuous index in a term, once as a subscript and once as a superscript, stand for integration with respect to that index over the fundamental interval* (a, b). *A parenthesis about a continuous index will denote an exception to this integration convention.*

A functional $\xi_\alpha[y^x]$ which under (1) transforms in accordance with the law

$$\bar{\xi}_\alpha[\bar{y}] = \xi_\alpha[y]_{,\alpha}y^{(\alpha)} + \xi_\sigma\,y^\sigma_{,\alpha} \tag{4}$$

will be called a covariant vector while a functional $\xi^\alpha[y^x]$ which under (1) transforms in accordance with the law

$$\bar{\xi}^\alpha[\bar{y}] = \xi^\alpha[y]_{,\alpha}\bar{y}^{(\alpha)} + \xi^\sigma\,\bar{y}^\alpha_{,\sigma} \tag{5}$$

a contravariant vector. A pair of functionals $t_{\alpha\beta}[y]$ and $t_\alpha[y]$ will be said to constitute a covariant tensor of rank two if under (1) it transforms in accordance with

$$\left.\begin{aligned}\bar{t}_{\alpha\beta}[\bar{y}] &= t_{\alpha\beta}[y]_{,\alpha}y^{(\alpha)}{}_{,\beta}y^{(\beta)} + t_{\sigma\beta}\,y^\sigma_{,\alpha}{}_{,\beta}y^{(\beta)} + t_{\alpha\sigma}\,y^\sigma_{,\beta}{}_{,\alpha}y^{(\alpha)} \\ &\quad + t_{\sigma\tau}\,y^\sigma_{,\alpha}\,y^\tau_{,\beta} + t_\alpha[y]y^\alpha_{,\beta}{}_{,\alpha}y^{(\alpha)} + t_\beta\,y^\beta_{,\alpha}{}_{,\beta}y^{(\beta)} + t_\sigma\,y^\sigma_{,\alpha}\,y^\sigma_{,\beta} \\ \bar{t}_\alpha[\bar{y}] &= ({}_{,\alpha}y^{(\alpha)})^2\,t_\alpha[y].\end{aligned}\right\} \tag{6}$$

Theorem I.—A necessary and sufficient condition that $\xi_\alpha[y]\delta y^\alpha$ be an absolute linear functional differential form is that ξ_α be a covariant functional vector. A necessary and sufficient condition that

$$t_{\alpha\beta}[y]\delta y^\alpha\,\delta y^\beta + t_\alpha(\delta y^\alpha)^2$$

be an absolute quadratic functional differential form is that the pair of functionals $t_{\alpha\beta}[y]$, $t_\alpha[y]$ form a covariant functional tensor of rank two.

Theorem II.—If

$$f_{\alpha\beta\ldots\gamma\delta}\, y^{\delta} + \varphi_{\alpha\beta\ldots\gamma}\, y^{(\alpha)} + \psi_{\alpha\beta\ldots\gamma}\, y^{(\beta)} + \ldots + \omega_{\alpha\beta\ldots\gamma}\, y^{(\gamma)} = 0,$$

for all continuous functions w^{α} such that $y^{a} = y^{b} = 0$, then in the field of continuous functions

$$f_{\alpha\beta\ldots\gamma\delta} \equiv 0, \quad \varphi_{\alpha\beta\ldots\gamma} \equiv 0, \quad \psi_{\alpha\beta\ldots\gamma} \equiv 0, \quad \ldots, \quad \omega_{\alpha\beta\ldots\gamma} \equiv 0.$$

Define the functional $Q^{i}[z^{\alpha}, w^{\alpha}]$ by

$$Q^{i}[z^{\alpha}, w^{\alpha}] = L^{i}_{\alpha\beta}[y] z^{\alpha} w^{\beta} + M^{i}_{\alpha}[y] z^{\alpha} w^{\alpha} + N^{i}_{\alpha}[y] z^{i} w^{\alpha} + O^{i}_{\alpha}[y] z^{\alpha} w^{j} + P^{i} z^{i} w^{j}.$$

Theorem III.—A necessary and sufficient condition that

$$Q^{i}[z^{\alpha}, w^{\alpha}] = Q^{i}[w^{\alpha}, z^{\alpha}]$$

for all continuous functions w^{α} and z^{α} which vanish at the end points of the fundamental interval (a, b) is that

$$L^{i}_{\alpha\beta} = L^{i}_{\beta\alpha}, \quad N^{i}_{\alpha} = O^{i}_{\alpha}.$$

The differentials of the functionals L, M, N, O, P will be assumed to be of the form

$$\left.\begin{aligned} \delta L^{i}_{\alpha\beta} &= {}_{\alpha,}L^{i}_{\alpha\beta}\, \delta y^{(\alpha)} + {}_{\beta,}L^{i}_{\alpha\beta}\, \delta y^{(\beta)} + {}_{i,}L^{i}_{\alpha\beta}\, \delta y^{(i)} + L^{i}_{\alpha\beta,\gamma}\, \delta y^{\gamma} \\ \delta M^{i}_{\alpha} &= {}_{\alpha,}M^{i}_{\alpha}\, \delta y^{(\alpha)} + {}_{i,}M^{i}_{\alpha}\, \delta y^{(i)} + M^{i}_{\alpha,\beta}\, \delta y^{\beta} \\ &\qquad\qquad \text{(similarly for } \delta N^{i}_{\alpha} \text{ and } \delta O^{i}_{\alpha}) \\ \delta P^{i} &= {}_{i,}P^{i}\, \delta y^{(i)} + P^{i}_{,\alpha}\, \delta y^{\alpha}. \end{aligned}\right\} \quad (7)$$

Theorem IV.—A necessary and sufficient condition that the functional (covariant functional differential)

$$\Delta\xi^{i} = \delta\xi^{i}[y] + L^{i}_{\alpha\beta}\, \xi^{\alpha}\, \delta y^{\beta} + M^{i}_{\alpha}\, \xi^{\alpha}\, \delta y^{\alpha} + N^{i}_{\alpha}\, \xi^{i}\, \delta y^{\alpha} + O^{i}_{\alpha}\, \xi^{\alpha}\, \delta y^{i} + P^{i}\, \xi^{i}\, \delta y^{j} \quad (8)$$

transform in accordance with the functional contravariant law (5) whenever $\xi^{i}[y]$ is a contravariant functional vector with a differential

$$\delta\xi^{i} = {}_{i,}\xi^{(i)}\, \delta y^{i} + \xi^{i}_{,\alpha}\, \delta y^{\alpha}$$

is that the functionals $L^{i}_{\alpha\beta}$, M^{i}_{α}, N^{i}_{α}, O^{i}_{α}, P^{i} have the following laws of transformation, under the functional transformations (1):

$$\left.\begin{aligned} \bar{L}^{i}_{jk}[\bar{y}] &= {}_{i,}\bar{y}^{i}\, \alpha^{(i)}_{jk} + \bar{y}^{i}_{,\sigma}\, \alpha^{\sigma}_{jk} + \bar{y}^{i}_{,j}\, \beta^{(j)}_{k} + \bar{y}^{i}_{,k}\, \gamma^{(k)}_{j} \\ \bar{M}^{i}_{j}[\bar{y}] &= {}_{i,}\bar{y}^{i}\, \epsilon^{(i)}_{j} + \bar{y}^{i}_{,\sigma}\, \epsilon^{\sigma}_{j} + \bar{y}^{i}_{,j}\, \zeta^{(j)} \\ \bar{N}^{i}_{j}[\bar{y}] &= {}_{i,}\bar{y}^{i}\, \beta^{(i)}_{j} \\ \bar{O}^{i}_{j}[\bar{y}] &= {}_{i,}\bar{y}^{i}\, \gamma^{(i)}_{j} \\ \bar{P}^{i}[\bar{y}] &= {}_{i,}\bar{y}^{i}\, \zeta^{(i)} \end{aligned}\right\} \quad (9)$$

The functionals α, β, γ, ϵ, ζ in (9) are defined by

$$\left.\begin{aligned}
\alpha^i_{jk} &= y^i_{,jk} + L^i_{jk\,j,}y^{(j)}{}_{k,}y^{(k)} + L^i_{j\sigma\,j,}y^{(j)}\,y^\sigma_{,k} + L^i_{\sigma k}\,y^\sigma_{j\,k,}y^{(k)} \\
&\quad + L^i_{\sigma\tau}\,y^\sigma_{,j}\,y^\tau_{,k} + M^i_{j\,j,}y^{(j)}\,y^j_{,k} + M^i_k\,y^{(k)}_{j\,k,}y^k + M^i_\sigma\,y^\sigma_{,j}\,y^\sigma_{,k} \\
&\quad + N^i_k\,y^i_{,j\,k,}y^{(k)} + N^i_\sigma\,y^i_{,j}\,y^\sigma_{,k} + O^i_{j\,j,}y^{(j)}\,y^i_{,k} + O^i_\sigma\,y^\sigma_{,j}\,y^i_{,k} \\
&\quad + P^i\,y^i_{,j}\,y^i_{,k} \\
\beta^j_k &= {}_{j,}y^{(j)}_{,k} + N^j_{k\,j,}y^{(j)}{}_{k,}y^{(k)} + N^j_{\sigma\,j,}y^{(j)}\,y^\sigma_{,k} + P^j{}_{j,}y^{(j)}\,y^j_{,k} \\
\gamma^j_k &= {}_{j,}y^{(j)}_{,k} + O^j_{k\,j,}y^{(j)}{}_{k,}y^{(k)} + O^j_{\sigma\,j,}y^{(j)}\,y^\sigma_{,k} + P^j{}_{j,}y^{(j)}\,y^j_{,k} \\
\epsilon^i_j &= {}_{j,}y^i_{,j} + M^i_j\,({}_{j,}y^{(j)})^2 \\
\zeta^i &= {}_{ii,}y^{(i)} + P^{(i)}\,({}_{i,}y^i)^2
\end{aligned}\right\} \tag{10}$$

For obvious reasons, the sequence of five functionals L^i_{jk}, M^i_j, N^i_j, O^i_j, P^i will be said to determine an *affine connection in our function space.* An affine connection for which

$$L^i_{jk} = \Gamma^i_{jk} \qquad (\Gamma^i_{jk} = \Gamma^i_{kj})$$
$$N^i_j = O^i_j$$

will be called a symmetric affine connection. It is clear from (9) that the conditions

$$\Gamma^i_{jk} = \Gamma^i_{kj}, \qquad N^i_j = O^i_j \tag{11}$$

are invariant under an arbitrary change of functional variable y^i. Thus a symmetric affine connection is determined by a sequence of four functionals Γ^i_{jk}, M^i_j, N^i_j, P^i whose law of transformation is obtained from (9) in an obvious manner. From now on we shall deal explicitly only with the symmetric affine connection. However, much that we shall have to say has corresponding parts in a theory of an asymmetric affine connection.

Let us define a pair of functionals $\xi^i_{;\alpha}$, ${}_{i;}\xi^{(i)}$ by

$$\left.\begin{aligned}
{}_{i;}\xi^{(i)} &= {}_{i,}\xi^{(i)} + \xi^\sigma\,N^i_\sigma + \xi^i\,P^i \\
\xi^i_{;\alpha} &= \xi^i_{,\alpha} + \xi^\sigma\,\Gamma^i_{\sigma\alpha} + \xi^{(\alpha)}\,M^i_\alpha + \xi^i\,N^i_\alpha.
\end{aligned}\right\} \tag{12}$$

The sequence ${}_{i;}\xi^{(i)}$, $\xi^i_{;\alpha}$ so defined will be called the *covariant functional derivative* of ξ^i with respect to the affine connection Γ, M, N, P. This definition is justified by the theorem.

Theorem V.—The covariant functional derivative of a contravariant functional vector $\xi^i[y]$ constitute a mixed functional tensor of rank two covariant of rank one and contravariant of rank one, i.e., the sequence of functionals ${}_{i;}\xi^{(i)}$, $\xi^i_{;\alpha}$ transforms in accordance with the law

$$\left.\begin{aligned}
{}_{i;}\bar{\xi}^{(i)} &= {}_{i;}\xi^{(i)} \\
\bar{\xi}^i_{;\alpha} &= \xi^i_{;\alpha}\,{}_{\alpha,}y^{(\alpha)}\,{}_{i,}\bar{y}^{(i)} + \xi^i_{;\sigma}\,y^\sigma_{,\alpha}\,{}_{i,}\bar{y}^{(i)} + \xi^\sigma_{;\alpha}\,{}_{\alpha,}y^{(\alpha)}\,y^i_{,\sigma} \\
&\quad + \xi^\sigma_{;\tau}\,y^\tau_{,\alpha}\,\bar{y}^i_{,\sigma} + {}_{i;}{}^{ii}{}_{,}\bar{y}^{(i)}\,y^i_{,\alpha} + {}_{\alpha;}\xi^\alpha\,{}_{\alpha,}y^{(\alpha)}\,\bar{y}^i_{,\alpha} + {}_{\sigma;}\,\xi^\sigma\,y^\sigma_{,\alpha}\,\bar{y}^i_{,\sigma}.
\end{aligned}\right\} \tag{13}$$

Clearly the *covariant functional differential* of ξ^i for the symmetric case can be written as

$$\Delta\xi^i = {}_{i;}\xi^i\,\delta y^{(i)} + \xi^i_{;\alpha}\,\delta y^\alpha. \tag{14}$$

Similar results hold good for the covariant functional derivative ${}_{i;}\xi_i$, $\xi_{i;\alpha}$ of the covariant functional vector ξ_i

$$\left.\begin{aligned} {}_{i;}\xi_i &= {}_{i,}\xi_i - \xi_\sigma M^\sigma_i - \xi_i P^{(i)} \\ \xi_{i;\alpha} &= \xi_{i,\alpha} - \xi_\alpha N^{(\alpha)}_i - \xi_i N^{(i)}_\alpha - \xi_\sigma \Gamma^\sigma_{i\alpha}. \end{aligned}\right\} \tag{15}$$

The functionals ${}_{i;}\xi_i$, $\xi_{i;\alpha}$ constitute a covariant functional tensor of rank two and hence transform in accordance with the law (6).

Theorem VI.—The conditions for the complete integrability of the variational equation[2]

$$\delta\xi^i + \Gamma^i_{\alpha\beta}\,\xi^\alpha\,\delta y^\beta + M^i_\alpha\,\xi^\alpha\,\delta y^\alpha + N^i_\alpha\,\xi^i\,\delta y^\alpha + N^i_\alpha\,\xi^\alpha\,\delta y^i + P^i\,\xi^i\,\delta y^i = 0 \tag{16}$$

are

$$B^i_{\alpha\beta\gamma} = 0, \quad C^i_{\alpha\beta} = 0, \quad D^i_{\alpha\beta} = 0, \quad E^i_\alpha = 0, \quad F^i_\alpha = 0 \tag{17}$$

where

$$\left.\begin{aligned} B^i_{\alpha\beta\gamma} &= \Gamma^i_{\alpha\beta,\gamma} - \Gamma_{\alpha\gamma,\beta} + \Gamma^\sigma_{\alpha\beta}\Gamma^i_{\sigma\gamma} - \Gamma^\sigma_{\alpha\gamma}\Gamma^i_{\sigma\beta} + \Gamma^i_{\alpha\beta}N^i_\gamma - \Gamma^i_{\alpha\gamma}N^i_\beta \\ &\quad + N^{(\alpha)}_\beta\Gamma^i_{\alpha\gamma} - N^{(\alpha)}_\gamma\Gamma^i_{\alpha\beta} + N^{(\beta)}_\alpha\Gamma^i_{\beta\gamma} - N^{(\gamma)}_\alpha\Gamma^i_{\gamma\beta} + \Gamma^{(\gamma)}_{\alpha\beta}M^i_\gamma \\ &\quad - \Gamma^{(\beta)}_{\alpha\gamma}M^i_\beta \\ C^i_{\alpha\beta} &= {}_{\alpha,}\Gamma^i_{\alpha\beta} - M^i_{\alpha,\beta} - M^\sigma_\alpha\Gamma^i_{\sigma\beta} - M^i_\alpha N^i_\beta + N^{(\alpha)}_\beta M^i_\alpha - P^{(\alpha)}\Gamma^i_{\alpha\beta} \\ &\quad - M^{(\beta)}_\alpha M^i_\beta \\ D^i_{\alpha\beta} &= {}_{i,}\Gamma^{(i)}_{\alpha\beta} - N^i_{\alpha,\beta} + \Gamma^\sigma_{\alpha\beta}N^i_\sigma + \Gamma^i_{\alpha\beta}P^i - N^i_\alpha N^i_\beta + N^{(\alpha)}_\beta N^i_\alpha \\ &\quad + N^{(\beta)}_\alpha N^i_\beta \\ E^i_\alpha &= {}_{i,}M^{(i)}_\alpha - {}_{\alpha,}N^i_\alpha + M^\sigma_\alpha N^i_\sigma + M^i_\alpha P^i + P^{(\alpha)}N^i_\alpha \\ F^i_\alpha &= {}_{i,}N^{(i)}_\alpha - P^i_{,\alpha}. \end{aligned}\right\} \tag{18}$$

Let $\xi^i[y]$, $\eta^i[y]$, $\zeta^i[y]$ be arbitrary contravariant functional vectors and $u_i[y]$ an arbitrary covariant functional vector. Then the general normal quadrilinear functional form $Q[\xi, \eta, \zeta, u]$ will have fifteen functional coefficients. The conditions

$$\left.\begin{aligned} Q[\xi, \eta, \zeta, u] &= -Q[\xi, \zeta, \eta, u] \\ Q[\xi, \eta, \zeta, u] + Q[\zeta, \xi, \eta, u] + Q[\eta, \zeta, \xi, u] &= 0 \end{aligned}\right\} \tag{19}$$

for arbitrary ξ, η, ζ, u cut down the number of independent functional coefficients to five. The functionals $B^i_{\alpha\beta\gamma}$, $C^i_{\alpha\beta}$, $D^i_{\alpha\beta}$, E^i_α, F^i_α, which constitute the *functional curvature tensor* based on a general symmetric affine connection of our function space, can be thought of as the functional coefficients of an absolute quadrilinear functional form $Q[\xi, \eta, \zeta, u]$ with symmetry conditions (19).

Definition of an absolute functional tensor.—A sequence of functionals of y^i that are the coefficients of an absolute multilinear functional form $M[\xi, \ldots, \eta, u, \ldots, v]$ in r contravariant vectors $\xi^i[y], \ldots, \eta^i[y]$ and s covariant vectors $u_i[y], \ldots, v_i[y]$ will be said to constitute an absolute functional tensor of rank $r + s$, covariant of rank r and contravariant of rank s.

The law of transformation of a functional tensor possesses the properties of linear homogeneity and transitivity.

A theory of covariant functional differentiation based on $\Gamma^i_{\alpha\beta}$, M^i_α, N^i_α, P^i is obtained by an extension of the elimination methods of my previous paper.[1] An alternative method is developed based on the covariant differential of the multilinear form M.

If we now assume that $y^x(t)$ is an analytic function of t for $t_0 \leq t \leq t_1$ and that $\Gamma^i_{\alpha\beta}$, M^i_α, N^i_α, P^i for an arbitrarily given $y^x(t)$ are analytic functions of t, it is possible to make the *integro-differential equation*

$$\left.\begin{aligned}&\frac{\partial y^j(t)}{\partial t}\left(\frac{\partial^2 y^i(t)}{\partial t^2} + \Gamma^i_{\alpha\beta}\frac{\partial y^\alpha}{\partial t}\frac{\partial y^\beta}{\partial t} + M^i_\alpha\left(\frac{\partial y^\alpha}{\partial t}\right)^2 + 2N^i_\alpha\frac{\partial y^i}{\partial t}\frac{\partial y^\alpha}{\partial t}\right.\\&\qquad\left. + P^i\left(\frac{\partial y^i}{\partial t}\right)^2\right)\\&= \frac{\partial y^i}{\partial t}\left(\frac{\partial^2 y^j(t)}{\partial t^2} + \Gamma^j_{\alpha\beta}\frac{\partial y^\alpha}{\partial t}\frac{\partial y^\beta}{\partial t} + M^j_\alpha\left(\frac{\partial y^\alpha}{\partial t}\right)^2 + 2N^j_\alpha\frac{\partial y^j}{\partial t}\frac{\partial y^\alpha}{\partial t}\right.\\&\qquad\left. + P^j\left(\frac{\partial y^j}{\partial t}\right)^2\right)\end{aligned}\right\} \quad (20)$$

the basis of a "geometry of paths" in our function space. For an "affine parameter" s defined along a functional path, the equations (20) for the path take the form

$$\frac{\partial^2 y^i(s)}{\partial s^2} + \Gamma^i_{\alpha\beta}\frac{\partial y^\alpha}{\partial s}\frac{\partial y^\beta}{\partial s} + M^i_\alpha\left(\frac{\partial y^\alpha}{\partial s}\right)^2 + 2N^i_\alpha\frac{\partial y^i}{\partial s}\frac{\partial y^\alpha}{\partial s} + P^i\left(\frac{\partial y^i}{\partial s}\right)^2 = 0. \quad (21)$$

For the theory of normal functional coördinates z^i we have the following theorem.

Theorem VII.—An arbitrary functional coördinate transformation

$$\bar{y}^i = \bar{y}^i[y]$$

induces a linear homogeneous functional transformation of the third kind

$$\bar{z}_i = (_{,i}\bar{y}^{(i)})_q\, z^i + (\bar{y}^i_{,\alpha})_q\, z^\alpha \qquad (D[\bar{y}^i_{,\alpha}/_{,\alpha}\bar{y}^{(\alpha)}])_q \neq 0$$

in the normal functional coördinates z^i, $\bar{z}^i$ with the same origin $y^i = q^i$.

The whole theory of the extensions of tensors (in particular the general

theory of covariant differentiation), normal tensors, replacement theorems, etc., of n dimensions have their analogues in our function space geometries. The proofs of the various general results are materially simplified by the employment of differentials of functionals rather than the functional derivatives, and the multilinear functional forms rather than the sequence of functionals that constitute the functional tensor.

A Riemannian function space geometry based on a functional quadratic differential form

$$ds^2 = g_{\alpha\beta}[y]\delta y^\alpha\,\delta y^\beta + g_\alpha[y](\delta y^\alpha)^2, \quad g_\alpha \neq 0, \quad g_{\alpha\beta} = g_{\beta\alpha}, \quad D[g_{\alpha\beta}/g_\alpha] \neq 0 \tag{22}$$

possesses geodesics that satisfy an integro-differential equation of the form (21).

Theorem VIII.—The covariant functional derivative $g_{\alpha\beta;\gamma}$, $_{\alpha;}g_{\alpha\beta}$, $_{\alpha;}g_\alpha$, $g_{\alpha;\beta}$ of the tensor $g_{\alpha\beta}$, g_α vanishes.

Corollary.—The functionals $g_{\alpha\beta}[y]$, $g_\alpha[y]$ are solutions of the system of variational equations

$$\delta g_{\alpha\beta} = \lambda_{\alpha\beta\gamma}[y]\delta y^\gamma + \mu_{\alpha\beta}[y]\,\delta y^{(\alpha)} + \mu_{\beta\alpha}[y]\,\delta y^{(\beta)}$$
$$\delta g_\alpha = \nu_{\alpha\beta}[y]\delta y^\beta + \rho_\alpha[y]\delta y^{(\alpha)},$$

where

$$\lambda_{\alpha\beta\gamma} = g_{\alpha\beta}N_\gamma^{(\alpha)} + g_{\alpha\beta}N_\gamma^{(\beta)} + g_{\gamma\beta}N_\alpha^{(\gamma)} + g_{\gamma\alpha}N_\beta^{(\gamma)} + g_{\sigma\beta}\Gamma_{\alpha\gamma}^\sigma + g_{\sigma\alpha}\Gamma_{\beta\gamma}^\sigma + g_\alpha\Gamma_{\beta\gamma}^{(\alpha)} + g_\beta\Gamma_{\alpha\gamma}^{(\beta)}$$

$$\mu_{\alpha\beta} = g_{\alpha\beta}P^{(\alpha)} + g_{\sigma\beta}M_\alpha^\sigma + g_\alpha N_\beta^{(\alpha)} + g_\beta M_\alpha^{(\beta)}$$

and $\Gamma_{\alpha\beta}^i$, M_α^i, N_α^i, P^i is the affine connection determined by (22).

Theorem IX.—The Fredholm determinant of $g_{\alpha\beta}/\sqrt{g_\alpha g_\beta}$ is a relative functional invariant of weight two of the quadratic functional form (22).

[1] A. D. Michal, *Am. J. Math.*, **40**, pp. 473–517, 1928. This paper was presented to the American Mathematical Society at the New York meeting, October, 1927. (Cf. *Bulletin of the Am. Math. Soc.*, **34**, pp. 8–9, 1928.)

[2] We have here the beginnings of a theory of variational equations that is so vital to the development of a group theory and invariant theory in function space.